UNIVERSITY*of*
TASMANIA

Uncertainty: types and applications in spatial predictive models

by

Majid Shadmanroodposhti

BSc. Geography, MSc. Remote Sensing and GIS

School of Technology, Environments and Design

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Geomatics Engineering

University of Tasmania

May 2019

Declaration of Originality

"This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright."

Majid Shadman

Jan 2019

Authority of Access
This thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act of 1968.

Publications included in this thesis
This thesis contains six published manuscripts in peer reviewed literature and have been incorporate as Chapters 2 to 7.

- **Chapter 2** → Shadman Roodposhti, M., Aryal, J., Lucieer, A. and Bryan, B.A., 2019. Uncertainty assessment of hyperspectral image classification: deep learning vs random forest. Entropy, 21(1): 78.

- **Chapter 3** → Shadman Roodposhti, M.; Lucieer, A.; Anees, A.; Bryan, B.A., 2019. A Robust Rule-Based Ensemble Framework Using Mean-Shift Segmentation for Hyperspectral Image Classification. Remote Sens. 11: 2057.

- **Chapter 4** → Shadman Roodposhti, M., Aryal, J. and Bryan, B.A., 2019. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. Environmental Modelling & Software, 112: 70-81.

- **Chapter 5** → Shadman Roodposhti, M., Safarrad, T. and Shahabi, H., 2017. Drought sensitivity mapping using two one-class support vector machine algorithms. Atmospheric Research, 193: 73-82.

- **Chapter 6** → Shadman Roodposhti, M., Aryal, J., Shahabi, H. and Safarrad, T., 2016. Fuzzy Shannon entropy: a hybrid GIS-based landslide susceptibility mapping method. Entropy, 18(10): 343.

- **Chapter 7** → Shadman Roodposhti, M., Hewitt, R.J. and Bryan, B.A., 2020. Towards automatic calibration of neighbourhood influence in cellular automata land-use models. Computers, Environment and Urban Systems, 79: 101416.

Statement of Co-authorship (where applicable)

The following people contributed to the publication of work undertaken as part of this thesis:

- Majid Shadman, School of Technology, Environment and Design, University of Tasmania
- Arko Lucieer, School of Technology, Environment and Design, University of Tasmania
- Jagannath Aryal, School of Technology, Environment and Design, University of Tasmania
- Brett A. Bryan, Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University
- Richard J. Hewitt, Information and Computational Sciences Group, James Hutton Institute, Aberdeen, Scotland, UK
- Asim Anees, Data Scientist Group, Children's Medical Research Institute, 214 Hawkesbury Road, Westmead, NSW. 2415 Australia
- Himan Shahabi, Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Iran
- Taher Safarrad, Department of Geography and Urban Planning, Faculty of Humanities and Social Science, University of Mazandaran, Iran

Author details and their roles:

- **Chapter 2:** Majid (85%), Jagannath (5%), Arko (5%) Brett (5%)

  All authors contributed to design and development. Majid performed all analysis and writing. Jagannath, Arko and Brett provided editorial advice.

- **Chapter 3:** Majid (85%), Arko (5%), Brett (5%), Asim (5%)

  All authors contributed to design and development. Majid performed all analysis and writing. Arko, Brett and Asim provided editorial advice.

- **Chapter 4:** Majid (85%), Jagannath (5%), Brett (10%)

  All authors contributed to design and development. Majid performed all analysis and writing. Brett majorly revised the write ups. Jagannath and Brett provided editorial advice.

- **Chapter 5:** Majid (90%), Himan (5%), Taher (5%)

  All authors contributed to design and development. Majid performed all analysis and writing. Himan and Taher provided editorial advice.

- **Chapter 6:** Majid (90%), Jagannath (4%), Himan (1%), Taher (5%)

  All authors contributed to design and development. Majid performed all analysis and writing. Jagannath, Himan and Taher provided editorial advice.

- **Chapter 7:** Majid (85%), Richard (12%), Brett (3%)

  All authors contributed to design and development. Majid performed all analysis and writing. Richard majorly revised the write ups. Richard and Brett provided editorial advice.

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above published peer-reviewed manuscripts contributing to this thesis:

Candidate: **Majid Shadman** 09/05/2019

Associate Professor Arko Lucieer Supervisor School of Technology, Environments and Design, University of Tasmania Date: 9 May 2019	**Professor Jason Byrne** Acting Head of School, Technology, Environments and Design, University of Tasmania Date: …… 17 Oct 2019

Abstract

Uncertainty is one of the most essential and fundamental issues that requires full attention in almost all spatial models and applications. Evidently, the quality of uncertainty modelling plays a critical role in resultant outcomes of geographical models and applications with an inevitable effect on decision- making processes. Therefore, up to now, uncertainty assessment and modelling has gained extensive attention in the field of spatial sciences. Considering the growing importance of this issue, this thesis investigates uncertainty modelling that applies in spatial science along with practical strategies to deal with them. To this end, three definitions of uncertainty are adopted, including Type A in which the uncertainties are derived from series of repeated observations, Type B , such as ambiguity and/or vagueness, with the uncertainties calculated by means other than the statistical analysis of series of observations (i.e. general knowledge of the behaviour and properties of phenomena) and Type C that is uncertainties in form of randomness. Proposed strategies to deal with each type of uncertainty is also exemplified in this thesis.

In terms of Type A uncertainty, in this thesis, entropy is the key term representing uncertainty. Considering the fact that repeated observations are more often aimed at implementation of predictive spatial models, three examples are described (1) assessing uncertainty of several machine-learning classification algorithms that are repeatedly applied in implementation of spatial predictive models, (2) improving performance of a classification scheme using uncertainty assessment and (3) applying an optimised algorithm using uncertainty assessment in land-use change simulation as a popular example. To this end, in Chapter 2, a strategy is proposed to evaluate the uncertainty of two machine-learning algorithms (i.e. random forest vs deep neural network) that are applied for land-use classification. These two algorithms are highly popular in implementation of predictive spatial models. In Chapter 3, a strategy to improve the performance of predictive algorithms using uncertainty models is proposed in the context of image classification. A comprehensive practical illustration is then applied for improving the quality of a land-use change model in Chapter 4.

With reference to Type B uncertainty, two different examples are implemented for drought sensitivity and landslide susceptibility mapping, where fuzzy If-Then rules and fuzzy sets are used to improve the quality of spatial models in Chapters 5 and 6, respectively.

Finally, Chapter 7 is focused on applying randomness for modelling Type C uncertainty that is frequently encountered in spatial models, especially land-use change models. Although randomness is an inevitable component of land-use change models, it is sometimes overlooked. In this chapter, randomness is applied as a primary component of land-use change models, but it is also applied to achieve the optimised neighbourhood setting as a key requirement of the model calibration phase.

This thesis, therefore, is a comprehensive illustration of different types of uncertainty in a range of important spatial models. Here, the thesis case studies are selected, so that they can be either applied to model the same spatial phenomenon within different case studies or can be generalised to new applications in the spatial domain.

This thesis is dedicated to:

God Almighty, for all the countless blessings and opportunities that he has bestowed upon me; I am dedicating this thesis to beloved people who have meant and continue to mean so much to me. My late mother and father for all their supports, prayers and wholehearted sacrifices full of love and devotion, they made in raising me and enabling me to achieve success in my life; My wife, Sepi who remains willing to engage with the struggle, and ensuing discomfort, of having a partner who refuses to give up his dreams. Without all of these contributions from these great personalities, I would probably not have been able to come this far in my educational career with this much success.

You are not a drop in the ocean. You are the entire ocean in a drop -Rumi

ACKNOWLEDGEMENT

The author would like to thank the following people / institutions for their help and valuable advices during this thesis:

- Tasmanian Graduate Research Scholarship (TGRS) and University of Tasmania, for funding this PhD thesis.
- Australian Sustainable Agriculture Scholarship (ASAS) and CSIRO for providing top-up scholarships and research funds.
- My supervisors, Arko Lucieer, Brett A Bryan and Jagannath Aryal for their support.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Overview

In this chapter, the intended meaning of uncertainty that applies to this thesis is defined and clarified. This is followed by a detailed description and categorisation of applied uncertainty models in spatial applications. Afterwards, the inherent necessity of modelling uncertainty as a prerequisite for modelling any spatial phenomena is explained in the problem statement. Finally, the aim and objectives of this thesis are outline, and the thesis structure is provided.

1.2. Uncertainty

The "*uncertainty*" phenomenon arises if there is a lack of information about the occurrence of some event. The fact that our descriptions of spatial phenomena are subject to uncertainty is now generally acknowledged. Spatial models are designed to handle large amounts of information about the natural and built environments where any collection of information observation is prone to uncertainty in a number of forms (Fisher, 1999). If that uncertainty is ignored, then there is always a chance for incorrect predictions or spatial simulation. Therefore, trust in the designed spatial model/application can be undermined. It is therefore of crucial importance to properly model uncertainty, in the best possible form, within a desired spatial model. Fundamental to that requirement is familiarity with popular forms of uncertainty that are repeatedly used for implementation of various spatial models (Figure 1.1) – the subject of this chapter.

Considering various types of spatial models, depending on the existing level of model complexity, one or more of the main uncertainty formats may be utilised (Figure 1.1). The aim of this chapter is to facilitate the conceptual understanding of the requirements for modelling different types of uncertainty within the spatial domain. Although no unanimous definition of uncertainty exists (Allard, 2013; Fisher, 1999; ISO and OIML, 1995; Taylor and Kuyatt, 1994), the most popular forms of uncertainty in applied spatial science can be divided in three major categories including uncertainties emanating from (1) *Type A* where the uncertainties is calculated from series of repeated observations such as those derived from probability distributions of desired estimates, (2) *Type B* such as ambiguity and/or vagueness where the uncertainty is modelled using experts' general knowledge about the behaviour and properties of phenomena (e.g. using fuzzy set and logic) (3) *Type C* where the uncertainty is an intermediate state of the first two categories (Figure 1.1).

The uncertainty component obtained from a Type A evaluation is calculated from repeated observations and is very similar to statistically estimated variance or the derived entropy from a set of probability values. A good example of Type A uncertainty that applies to a geographical phenomenon is determination of a land-use class for a given pixel in a satellite image. The concept of land-use is fundamental where the definition of each land-use class is defined by a detailed standard. In this regard, for the purpose of land-use classification/prediction, the available attribute options to be assigned to a land-use pixel or parcel is limited but uncertain. Here, the probability of each possible outcome can be quantified using land-use classification/prediction algorithms. Thus, the interpretation of the uncertainty concept is usually uncomplicated in its spatial expression (i.e. using observed frequency distribution).

Type B evaluation of uncertainty is to deal with ambiguity and/or vagueness defined as the method of evaluation of uncertainty by means of existing general knowledge or experts' knowledge. An example for this uncertainty type is the definition of a class (i.e. steep slope) or set (i.e. suitable or unsuitable pixel values) is a matter of vagueness, which can be treated by fuzzy set theory. Many examples of a poorly defined geographical phenomenon are applicable to standardisation of spatial criteria in the process of suitability or susceptibility mapping (Chang et al., 2008; Charabi and Gastli, 2011; Feizizadeh et al., 2014c; Malczewski, 2006; Shadman Roodposhti et al., 2016). For instance, the definition of a steep slope is often vague where the intermediate border between level slope, gentle slope and steep slope are often gradual and non-specific.

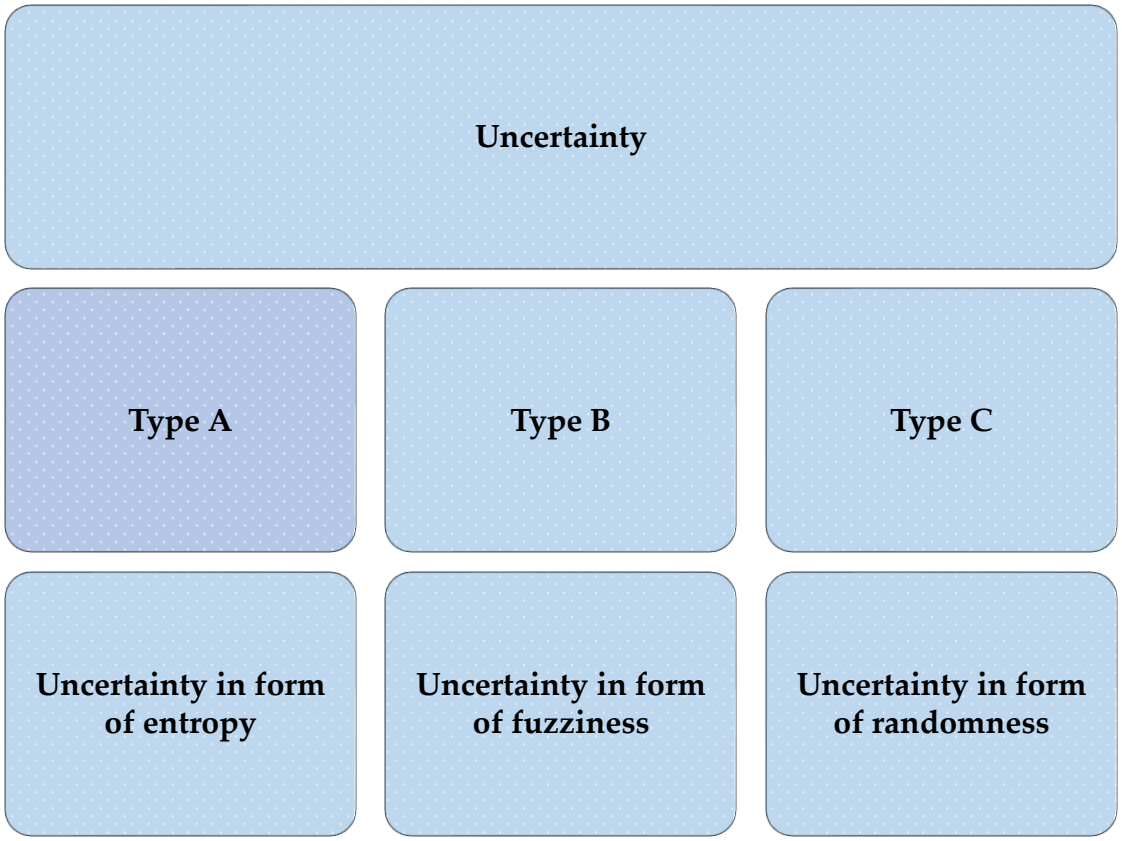


Figure. 1.1. Popular forms of uncertainty that are repeatedly used for implementation of various spatial models (adapted from (ISO and OIML, 1995), with revisions).

In this thesis, Type C uncertainty is then identified as an intermediate state between the first two uncertainty types. Although there are similarities between random effects and Type A uncertainty and also between systematic effects and Type B uncertainty, they are not the same. Random effects are usually categorised as Type A; however, they also can be evaluated by Type B estimates, especially when the magnitude of the randomness effect is available without performing repeated measurements. For instance, in land-use change models, the predicted transitions are implemented by comparing the class-specific probabilities for each pixel to a random number chosen from a Uniform, Weibull, Poisson distribution etc. between 0 and 1. If the scaled transition probability to a new land-use class matches the random value, the transition takes place; otherwise, the grid cell maintains its current land cover (Hepinstall et al., 2008). This is to compensate for existing levels of uncertainty in LUC models while applying individual measurements to calibrate each desired random distribution (i.e. Weibull distribution), which is often omitted.

Ignoring uncertainty undermines trust in the designed spatial model/application and corresponding decision-making phase; therefore, it is of crucial importance to properly model uncertainty, in the best possible form, within a desired spatial model. Obviously, fundamental to that requirement is familiarity with popular forms of uncertainty that are repeatedly used for implementation of various spatial models. To this end, applications that are discussed in this thesis are centralised around uncertainties derived from probability distribution, fuzzy set and logic and randomness as representative of Types A, B and C uncertainties. Posterior to defining three types of uncertainty that exist in spatial models, each definition is followed at least with one detailed application. This will not only respond to the gap of defining uncertainty in spatial models, but also further facilitate the conceptual understanding of the requirements for modelling different types of uncertainty within the spatial applications.

1.2.1. Uncertainty in form of entropy

Probability is a study of the frequency of occurrence in crisp events such as the problem of quantifying the probability of correct classification or assigning a correct class label to an image object. In spatial science, the assessment of probability related uncertainties in the form of entropy, information gain, Gini impurity etc. is usually aimed to improve the quality of classification/prediction -as in deep neural networks (Guo et al., 2017; LeCun et al., 2015)- or mapping probability of correct classification/prediction -as an uncertainty assessment approach (Khatami et al., 2017). Although a variety of methods can be applied for this purposes, using entropy as the quantitative measure of system disorder or instability (Shannon, 2001) is more frequent throughout contemporary literature (da Silva et al., 2016; He and Kolovos, 2018; Porwollik et al., 2017).

1.2.2. Uncertainty in form of fuzziness

Vagueness arises when we are planning to group some objects that have a common property φ. The outcome is an actualized grouping of objects, X, which is not necessarily a set because the property φ may be vague. In simple words, it may not be always possible to characterise all the elements of the given grouping precisely and unambiguously. The fuzzy set theory offers a well-established mathematical framework in which conceptually vague phenomena can be precisely and rigorously studied (Zimmermann, 2010). Fuzzy sets can trace and quantify uncertainties within the data efficiently overriding intrinsic subjectivity within the data domain by means of fuzzification, where crisp variables are altered into fuzzy or continuous inputs (Srivastava et al., 2016; Zadeh, 1965a; Zadeh, 1976). The fuzzy membership function (FMF) is a generalisation of the indicator function in classical sets. In fuzzy logic, it represents the degree of truth as an extension of valuation. Degrees of truth are often confused with probabilities, although they are conceptually distinct, because fuzzy truth represents membership in vaguely defined sets, rather than the likelihood of some event or condition. FMF were introduced by Zadeh (1965a) who proposed using a membership function (with a range covering the interval (0,1)) operating on the domain of all possible values. In terms of spatial applications, vagueness is considered by applying FMF for data standardisation (Donevska et al., 2012; Feizizadeh et al., 2014c; Makropoulos et al., 2003; Roodposhti et al., 2014a) in various applications.

Another common form of modelling vagueness in spatial applications is through the implementation of fuzzy rules (If-Then) rules, such as those forming fuzzy inference systems (FIS). Fuzzy rules are used within fuzzy systems to infer an outcome based on the input variables (Enderton and Enderton, 2001). These if-then rule statements are used to formulate the conditional statements that comprise fuzzy logic (Roodposhti et al., 2017).

1.2.3. Uncertainty in form of Randomness

Randomness is a good representative of ambiguity that is highly applied in spatial applications especially land-use simulation models. In statistics, a random variable is an assignment of a numerical value to each possible outcome of an event space. This association facilitates the identification and the calculation of probabilities of the events. Randomness is the lack of pattern or predictability in events. A random sequence of events, symbols or steps has no order and does not follow an intelligible pattern or combination.

1.3. Modelling uncertainty in spatial applications

Modelling uncertainty of geographic objects/processes has been long considered throughout the modern spatial literature. The literature on uncertainty includes many definitions and applications, with various levels of complications. Here, some of the most popular examples of highly cited research papers for modelling uncertainty of geographic objects/processes are listed within the modern spatial literature (Table 1.1).

Table 1.1. Examples of highly cited research papers for modelling uncertainty of geographic objects/processes throughout the modern spatial literature

	Application	Reference
Probability	Prioritising spatial criteria for landslide susceptibility mapping	(Devkota et al., 2013b; Felicísimo et al., 2013; Pourghasemi et al., 2012a)
	Prioritising spatial criteria for site selection studies	(San Cristóbal, 2011)
	Environmental vulnerability assessment	
	Feature selection for classification and prediction tasks	(Löw et al., 2013)
	Prioritising spatial criteria for multi-criteria decision analysis	(Jaafari et al., 2014; Pourghasemi et al., 2012a)

	Application	Reference
Fuzzy set and logic	Standardisation of landslide susceptibility mapping criteria using FMF	(Akgun and Türk, 2010a; Feizizadeh et al., 2014c; Pradhan, 2011)
	Landslide susceptibility mapping using If-Then rules	(Akgun et al., 2012b; Pourghasemi et al., 2012b)
	Standardisation of spatial criteria for site selection using FMF	(Aydin et al., 2013; Gbanie et al., 2013; Moeinaddini et al., 2010)
	Prioritising spatial criteria for flood susceptibility mapping	(Tehrany et al., 2014; Zou et al., 2013)
	Prioritising spatial criteria for multi-criteria decision analysis	(Aydin et al., 2010; Guo and Zhao, 2015; Kaya and Kahraman, 2010)

	Application	Reference
Randomness	Modelling land-use change using cellular automata/agent-based models	(Aguilera et al., 2011; Valbuena et al., 2010)
	Modelling traffic density	(Lee et al., 2014)

Each publication above has minimum citation of a hundred where the publication date is no older than 2010.

1.4. Problem statement

Modelling uncertainty is a prerequisites for model building in any field where models are used, including spatial science (Crosetto et al., 2000). The majority of spatial data and models contain numerous inherent and introduced types of uncertainty (Davis and Keller, 1997). To date, research into uncertainty management has focused on identification and management of the different individual types of uncertainties and how they can be represented in simple overlay procedures. However, addressing uncertainty is an ongoing creative exploration (Li et al., 2018b) and yet the comprehensive categorisation of major types of uncertainty in spatial modelling and simulation has not been updated by applying state-of-the-art methodologies. Especially in the era of big geospatial data, spatial analyses and spatial datasets evolve with technological advances; therefore, new methods for studying uncertainty will be required in spatial science. These methods shall be contributing to quantifying predictive uncertainty of emerging spatial models that is a challenging and yet unsolved problem. Understanding the influence of existing uncertainties in the quality of spatial modelling and simulation

for various environmental applications requires in-depth investigation of diverse types and levels of uncertainties within spatial data/models. It is thus critical that spatial modellers know the importance of quantifying existing uncertainties in their typical forms. Thus, this thesis investigates suitable approaches for modelling each existing form of spatial uncertainties including: Type A, B and C that are applicable to popular spatial applications namely image processing, land-use change modelling, sensitivity and susceptibility mapping.

Modelling uncertainty allows us to assess the uncertainty that is linked with the model response because of uncertainties in the model input (Crosetto and Tarantola, 2001; Crosetto et al., 2000). Therefore, this thesis aims to investigate whether modelling uncertainty can assist in improving the quality of existing spatial applications in the context of land-use change modelling, remote sensing image classification (Martin et al., 2006; Weng et al., 2018), suitability (Ray and Burgman, 2006), sensitivity (Johnson and Gillingham, 2004) and susceptibility mapping (Shadman Roodposhti et al., 2016), which are among most popular spatial models in contemporary literature. or instance, numerous machine-learning and statistical methods have been used to calculate land-use transition rules and map transition potential for use in cellular automata land-use models (Roodposhti et al., 2018). Each of these methods employs structurally different numerical formulations, which affect the accuracy and transparency of LUC models. However, few of these methods facilitate the transparent extraction of transition rules and their corresponding uncertainty. A similar problem also exists in terms of detecting classification rules in image processing, outlining mechanisms in landslide susceptibility mapping, etc. In terms of the landslide susceptibility mapping or drought sensitivity mapping, lack of model transparency and inefficiencies in modelling uncertainty is more tangible as it will directly affect our interpretation of the model outcome. This is again elevated when it comes to state-of-the-art spatial models with higher levels of complexity. To contribute in solving this issue, six individual case studies for different modelling applications are implemented in this thesis. The proposed methods and workflows can handle models of a high degree of complexity and can also be generalised to other environmental applications in the spatial domain.

1.5. Aim and objectives

The overarching aim of this thesis is to assess and implement novel procedures to deal with major types of uncertainty in spatial models. The research is designed to improve the quality of decision-making based on implemented spatial models. To achieve the overarching aim of this research, the following objectives have been defined:

(1) To develop and compare an uncertainty assessment technique as a spatial approximator of classification accuracy, which can be used to identify unreliable pixel-level class allocations (with high uncertainty).

(2) To investigate uncertainty assessment that is derived from the probability distribution of possible outcomes of hyperspectral image classification. This is implemented by a highly accurate and transparent rule-based image classification algorithm.

(3) To apply uncertainty assessment derived from the probability distribution of land-use for implementation of transparent transition potential maps, which is applied for modelling land-use change simulation.

(4) To implement a novel drought sensitivity mapping technique through modelling uncertainty levels of decision rules using two one-class support vector machines (OC-SVM).

(5) To develop an optimal solution to deal with spatial data uncertainty for both criteria standardisation and prioritisation that is applicable to susceptibility mapping. This objective is

especially highly beneficial when there is not sufficient knowledge about a set of spatial criteria belonging to a desired area of interest.

(6) To assess the possibilities of applying randomness for better calibration of neighbourhood analysis required in automata land-use change models.

1.6. Thesis Outline

This thesis is composed of published journal manuscripts and manuscripts currently under review. Chapters 2, 4, 5 and 6 have been published in peer-refereed journals. Chapter 3 and 7 have been submitted to scientific journals for publication. The thesis structure diagram (Figure 1.2) shows how the three types of uncertainty relate to the individual chapters and thesis objectives.

Chapters 2 to 7 present specific case studies to exemplify different types of uncertainty (i.e. Types A, B and C). In terms of Type A uncertainty, three examples are described in a sequential order, where entropy is used as a key concept representing this type of uncertainty. Chapter 2, is aimed at assessing uncertainty of random forest and deep neural network as popular machine-learning algorithms that are repeatedly applied to implement a variety of spatial predictive models. Then, chapter 3 is an attempt to improve the performance of a novel rule-based ensemble classification scheme using uncertainty assessment and findings of chapter 2. Afterwards, chapter 4 is a practical example of applying an optimised algorithm using uncertainty assessment (i.e. such as those applied in chapter 3) in the context of land-use change simulation as a popular example in spatial modelling literature. Chapters 5 and 6 are dedicated to Type B uncertainty, where uncertainty is applied to deal with the vagueness of spatial models. Thus, two different examples are implemented for drought sensitivity and landslide susceptibility mapping, where fuzzy If-Then rules and fuzzy sets are applied to improve the quality of spatial models in Chapters 5 and 6, respectively. Finally, Chapter 7 is dedicated to randomness as an example of modelling uncertainty Type C that is frequently encountered in spatial models, especially land-use change models. Although randomness is an inevitable component of land-use change models, it is sometimes overlooked. In this chapter, randomness is applied as a primary component of land-use change models. In addition, it is demonstrated that randomness can also be utilised to achieve an optimised neighbourhood setting as a key requirement of the model calibration phase.

1.7. Future research and application to other datasets

The proposed methodologies for modelling various aspects of uncertainty in this thesis are not limited to specific geographic location and any similar datasets can be utilised in any potential future research. In this regard, those datasets applied in chapters 2 and 3 (i.e. Salinas dataset, Indian Pines dataset and Pavia University dataset) are publicly available world reference datasets that have been repeatedly used by other studies for testing various novel methodologies applicable to hyperspectral imagery. All three datasets that are used in Chapters 2 and 3 are available at http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. The datasets that are applied for methodology implementation of chapters 4, 5, 6 and 7 belong to Iran, which are applied with respect to the fact that there were prior knowledge and proper understanding about the datasets involved in the implementation of spatial models in each chapter. For every chapter (i.e. chapters 2 to 7), where research objective applies to Australian case studies, there will be no limit to benefit from Australian datasets.

As a matter of fact, all the introduced modelling approaches in this thesis are numeric and quantitative models that may be applied especially to similar topics. Some can be also applied to other global contexts. For example, applying uncertainty measures (e.g. Shannon entropy or any other form of uncertainty) for assessing predictivity strength (e.g. chapter 2 and 3), prioritisation of predictor variables (e.g. chapter 4), determining the weight of predictor variables or spatial criteria in variety of spatial multi-criteria models (such as chapter 6) does not belong to specific geographic boundaries and can be beneficial for problem-solving purposes without spatial or temporal limits. This also applies to

fuzzy rules (e.g. If-Then rules) to deal with spatial uncertainties. For better demonstration purpose a sample work for Australian case study is prepared (Roodposhti et al., 2019b), where a new methodology is developed by combining methodological contents of Chapters 2 to 6 and applied for the purpose of landslide susceptibility mapping where Shannon Entropy is used for criteria prioritisation and a set of fuzzy If-Then rules are applied to convert susceptibility scores to susceptibility classes with minimum possible subjectivity.

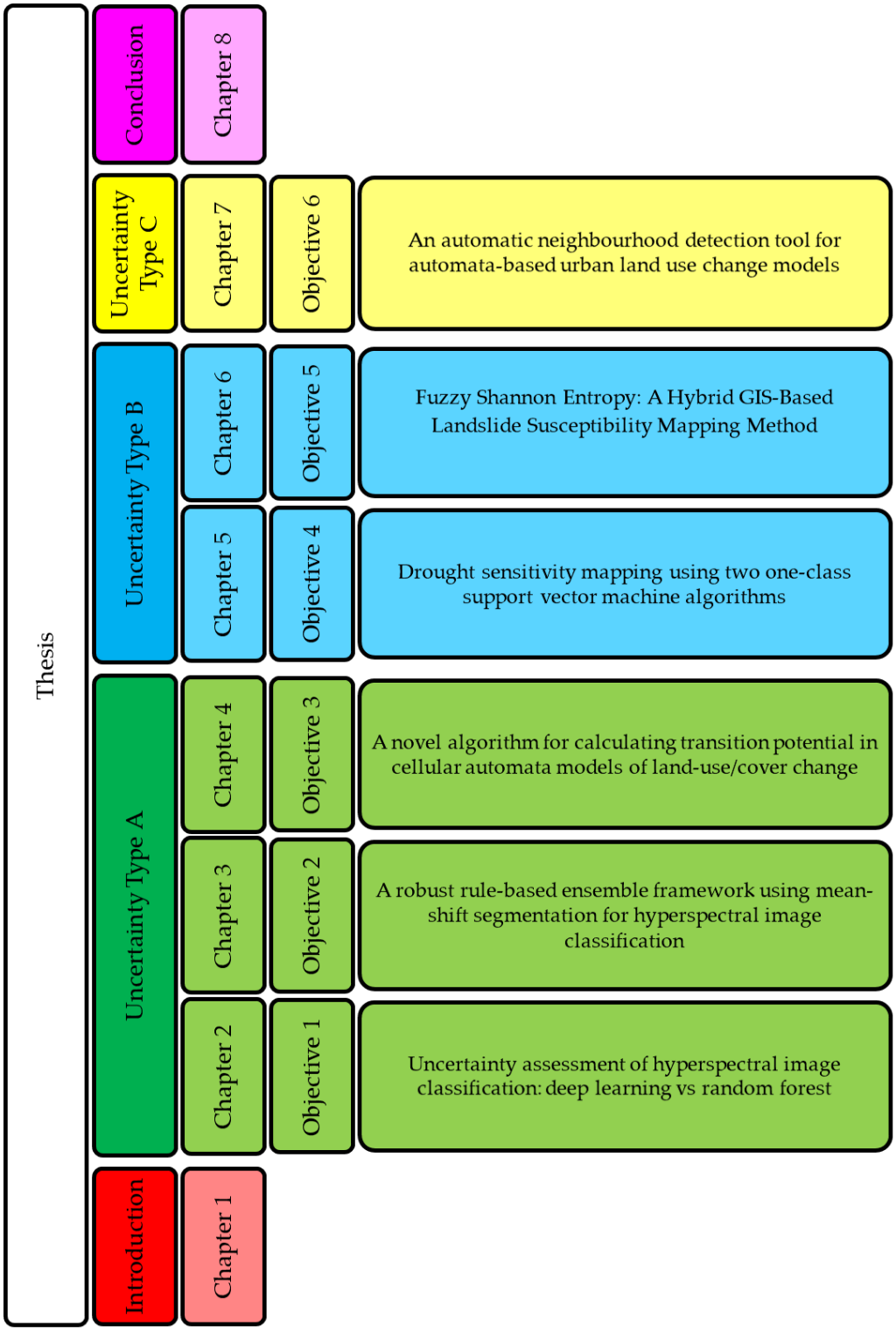


Figure. 1.2. Structure of thesis and achievement of objectives.

2. Uncertainty assessment of hyperspectral image classification: deep learning vs random forest

The current chapter is aimed to apply and compare two uncertainty assessment techniques that do not rely on test data availability and enable the spatial characterisation of classification accuracy before the validation phase, promoting the assessment of error propagation within the classified imagery products. This chapter is an attempt to model uncertainty Type A and has been published in the Entropy journal.

Shadman Roodposhti, M., Aryal, J., Lucieer, A. and Bryan, B.A., 2019. Uncertainty assessment of hyperspectral image classification: deep learning vs random forest. Entropy, 21(1): 78.

Abstract: Uncertainty assessment techniques have been extensively applied as an estimate of accuracy to compensate for weaknesses with traditional approaches. Traditional approaches to map accuracy assessment have been based on a confusion matrix and hence, are not only dependent on the availability of test data but are also incapable of capturing the spatial variation in classification error. Here, we apply and compare two uncertainty assessment techniques that do not rely on test data availability and enable the spatial characterisation of classification accuracy before the validation phase, promoting the assessment of error propagation within the classified imagery products. We compared the performance of emerging deep neural network (DNN) with the popular random forest (RF) technique. Uncertainty assessment was implemented by calculating the Shannon entropy of class probabilities predicted by DNN and RF for every pixel. The classification uncertainty of DNN and RF is quantified for two different hyperspectral image datasets—Salinas and Indian Pines. We then compared the uncertainty against the classification accuracy of the techniques represented by a modified root mean square error (RMSE). The results indicate that considering modified RMSE values for various sample sizes of both datasets, the derived entropy based on the DNN algorithm was a better estimate of classification accuracy and hence, provided a superior uncertainty estimate at pixel level.

2.1. Introduction

Assessing and mapping the state of the Earth's surface is a key requirement for many global research in the context of natural resources management (Xie et al., 2008), natural hazards modelling (Dutta et al., 2016; Roodposhti et al., 2017), urban planning (Weng et al., 2018; Xiao et al., 2006) etc., where all these mapping products need to be validated (Congalton, 1991; Congalton and Green, 2008). With the initiation of more advanced digital satellite remote sensing techniques, accuracy assessment of emerging methods has received major interest (Congalton and Green, 2008). The conventional way to report classification and/or prediction map accuracy is through an error matrix estimated from a test dataset, which is independent of the training process (Khatami et al., 2017). Accuracy metrics such as Cohen's Kappa coefficient (Cohen, 1968), overall accuracy (Congalton, 1991) and class-specific measures such as user's and producer's accuracies are usually estimated based on an error matrix (Richards and Jia, 1999b). However, it is not clear how these accuracy metrics relate to per-pixel accuracy (Ye et al., 2018) as these types of accuracy metrics are incapable of understanding the spatial variation of classification accuracies despite its importance in modelling spatial phenomena (Comber et al., 2012; Foody, 2002).

Different approaches have been proposed to characterize the quality of classified maps at the local scale (Khatami et al., 2017). One method is to apply empirical models to link classification accuracy (dependent variable) to different independent (predictor) variables, such as land-cover class (Burnicki, 2011; Yu et al., 2008). As the dependent variable is dichotomous (i.e., classified correctly or not), logistic regression is the most frequently applied algorithm for this purpose. Another approach for characterizing map quality at the local scale involves spatial interpolation of classification accuracy of the test dataset (Tsutsumida and Comber, 2015). The most recent approach is introduced by Khatami et al (2017), building on Stehman (1997). Here, a per-pixel accuracy prediction is implemented by applying different accuracy prediction methods based on four factors including predictive domain (spatial or spectral), interpolation function (constant, linear, Gaussian, and logistic), incorporation of

class information (interpolating each class separately versus grouping them together), and sample size. The fourth and most popular approach (Khatami et al., 2017) is to use the probabilities of class memberships or prediction strength (i.e. tree votes in the random forest or probabilities in neural networks) as indicators of classification uncertainty. The idea is that for a certain pixel, the greater the probability of class membership for a given labelled class, the lower the uncertainty associated with that class and analytical functions can be used to quantify the uncertainty measures instead of using only the membership value of the most probable class. Examples of these functions include ignorance uncertainty (Legleiter and Goodchild, 2005), Shannon's entropy (Dehghan and Ghassemian, 2006; Loosvelt et al., 2012), and α-quadratic entropy and maximum probability (Giacco et al., 2010), where entropy summarizes the information from membership values of all classes.

Uncertainty assessment techniques can provide an uncertainty map as a spatial approximator of classification accuracy, which can be used to locate and segregate unreliable pixel-level class allocations from reliable ones. In addition, this approach is independent of test data availability. This uncertainty assessment may be implemented using two types of classification approaches: unsupervised schemes using no training dataset (Prasad and Arora, 2014; Wang and Shi, 2013), and supervised schemes (Brown et al., 2009; Foody et al., 1992; Loosvelt et al., 2012; McIver and Friedl, 2001). Although unsupervised approaches can be applied regardless of the training dataset availability (i.e. by applying unsupervised algorithms), their relevant uncertainty assessment results may be misleading due to incorrect classification of pixels. In terms of supervised methods, various algorithms have been applied to evaluate the uncertainty of correct/incorrect classified pixels including random forest (RF) as one of the most popular algorithms. RF (Breiman, 1996; Breiman, 2001) has a rich and successful history in machine learning including applications in hyperspectral image classification (Abdel-Rahman et al., 2015; Chan and Paelinckx, 2008; Crawford et al., 2003; Lawrence et al., 2006; Naidoo et al., 2012) and uncertainty assessment (Coulston et al., 2016; Ließ et al., 2012; Loosvelt et al., 2014). It has been demonstrated to outperform most state-of-the-art learners when it comes to handling high dimensional data (Caruana et al., 2008), such as hyperspectral image datasets. Nonetheless, we assumed that considering high-dimensional hyperspectral data, newly emerging deep learning algorithms may be efficient for uncertainty assessment, but they have been rarely applied for this purpose. On the other hand, the deep learning algorithms have also been found to be more accurate than traditional algorithms, especially for image classification (Chen et al., 2014; Hinton and Salakhutdinov, 2006; Krizhevsky et al., 2012). Further, with multiple layers of processing, they may extract more abstract, invariant features of data, which is considered beneficial for uncertainty assessment studies.

Uncertainty assessment techniques have been repeatedly applied to assess the quality of hyperspectral image classification (Acquarelli et al., 2017; Adep et al., 2016; Wang and Shi, 2013). While deep learning has attracted broad attention as a classification algorithm (Liu et al., 2017; Pan et al., 2017; Tao et al., 2018; Wei et al., 2018) it has not been applied to uncertainty assessment of hyperspectral image classification nor compared to other methods. Thus, here we aim to apply DNN for uncertainty assessment of correct/incorrect classification for every pixel and then compare it with RF. Due to its high performance in uncertainty assessment studies, the RF algorithm provides an appropriate benchmark for comparing the performance of uncertainty assessment derived from deep learning. This paper aims to explore, quantify and compare the capability of DNN and RF algorithms for uncertainty assessment of hyperspectral imagery using two different hyperspectral datasets. To this end, by applying DNN in this study, we compare the uncertainty assessment of hyperspectral image classification using probability values derived from deep learning neurons and popularity votes of random forest (RF) trees combined with uncertainty values using Shannon entropy.

2.2. Methods and dataset

2.2.1. Method

This study follows two major steps (Figure 2.1). In step 1, the whole dataset is randomly divided into training (50%) and test data (50%). For each dataset, the hyper-parameters of the optimum DNN

and RF algorithms (Table 2.1) are configured using a 5-fold cross-validation of the training data in pre-processing stage. This is done only for hyper-parameters with a significant effect on the datasets and the remaining hyper-parameters are kept at the default values. Although test data is always a constant sub-set of the whole dataset, training procedure is done using different portions of training sample (i.e. 10%, 20%, … 100%) to assess the effects of training sample size in uncertainty assessment. Thus, the training sample itself is sliced into 10 equal random portions, and then applied for training the tuned algorithms. The algorithms are then trained 10 times each, from 10% to 100%, every time by a 10% increase of training samples, x={10%, 20%, …, 100%}, while the test dataset is always the same. In addition, to achieve more consistent results and to account for sensitivity analysis, each algorithm is applied in five consecutive runs, where the sampling strategy is the same but with location of initial sampling seeds (i.e. random training (50%) and test data (50%)) is modified by a different random function. As the hyper-parameters of the DNN and RF algorithms are optimised using a validation sample, they were not modified for the other sample sizes. Here, for both DNN and RF, the probability of belonging to each possible class was estimated for every pixel and used to compute the uncertainty of classification for that pixel using Shannon's entropy (Dehghan and Ghassemian, 2006), where entropy represents uncertainty in this research (Khatami et al., 2017).

Table 2.1. The optimised hyper-parameters of DNN and RF using 5-fold cross-validation data for uncertainty assessment.

Algorithm	Hyper-parameter	Description	Salinas	Indian Pines
DNN	hidden	Hidden layer sizes	[100,100]	[200, 200]
DNN	epoch	How many times the dataset should be iterated (streamed)	300	300
DNN	activation	Activation function for non-linear transformation.	"Maxout"	"Maxout"
DNN	stopping metric	A metric that is used as stopping criterion	"RMSE"	"RMSE"
DNN	l1	Only allows strong value to survives	0.0001	0.0001
DNN	l2	Prevents any single weight from getting too big	0.001	0.001
DNN	epsilon	Prevents getting stuck in local optima	1e-10	1e-10
RF	ntree	Number of trees to grow	100	100
RF	mtry	Number of variables available for splitting at each tree node	14	15

*For deep learning this optimisation is done using "Grid Search" by h20.grid() function and for Random forest it has been done manually for number of trees while tunerf() function is used to optimise mtry.

In step 2, for a better demonstration of classification performance considering the low and high uncertainty values, we mapped the uncertainty outputs along with the mode of correct/incorrect classified test pixels for all applied training samples (i.e. from 10% to 100%). Whenever an optimised algorithm is applied in the context of uncertainty assessment, the uncertainty value for a correctly classified pixel should be minimised (i.e. "0") while it should be maximised (i.e. "1") for misclassified pixels. Thus, we then calculated root mean square error (RMSE) of every prediction implemented by each algorithm (Dehghan and Ghassemian, 2006) to quantify the degree of deviation from this optimum state. For this purpose entropy values are normalised between 0 and 1. This whole process was implemented in R (R Core Team, 2017) using three major packages namely "H2O" (Candel et al., 2016), "randomforest" (Liaw and Wiener, 2002), "entropy" (Hausser et al., 2012).

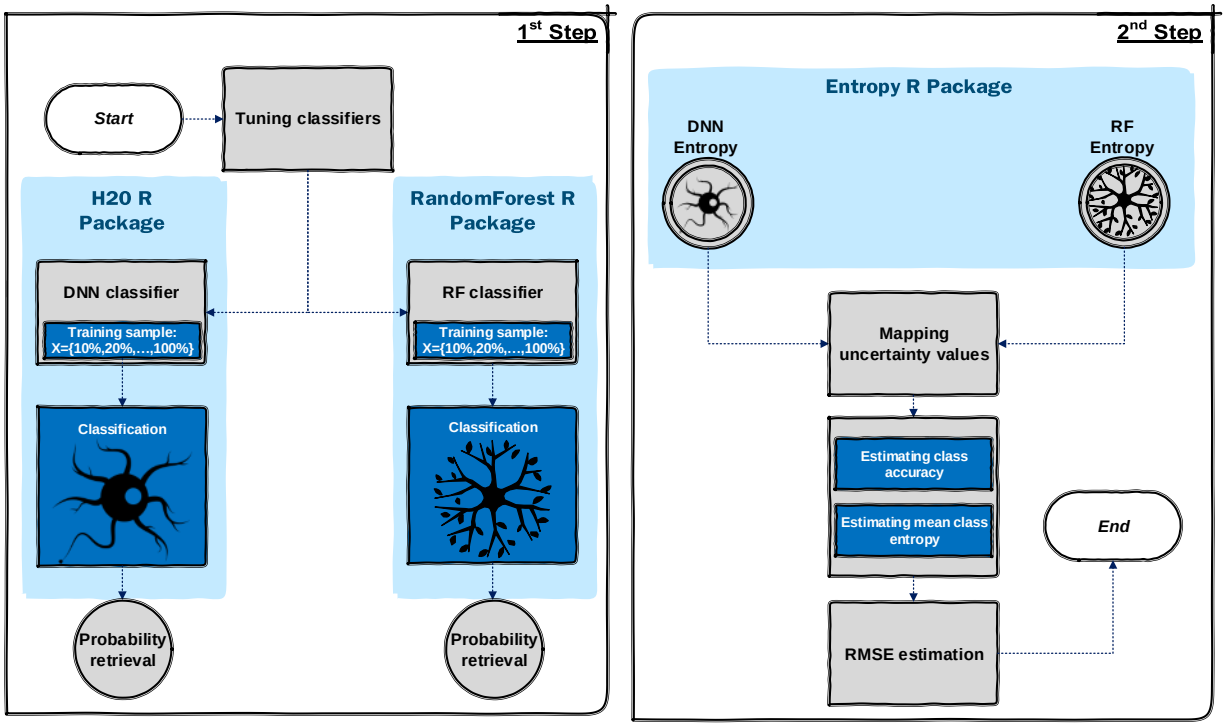


Figure 2.1. Flowchart of methodology implementation labelled with the main R packages utilized.

2.2.1.1. Supervised uncertainty assessment approach

The most popular and accurate way of uncertainty assessment is based on a supervised scheme using a machine-learning algorithm. Here, we implemented a model that can assess uncertainty values of a classified hyperspectral image containing various class labels. We first collected ground truth data labelled with their class categories such as corn, grass, hay, oats, soybean, etc. During training, the algorithm is provided with a training example and produces a response in the form of a vector of probabilities, one for each class. Then, the best-case scenario would be the highest probability score for one class and the lowest possible probability score for other remaining classes. The least desirable case, on the other hand, would be equal probability scores for all the existing class labels (Figure 2.2). We then compute the uncertainty of probability scores for all potential class labels for a pixel by using entropy. An ideal machine-learning algorithm, for uncertainty assessment, is not only capable of classifying input data with the highest possible accuracy but also it is also capable of producing class labels with low uncertainty for correctly classified pixels and vice versa.

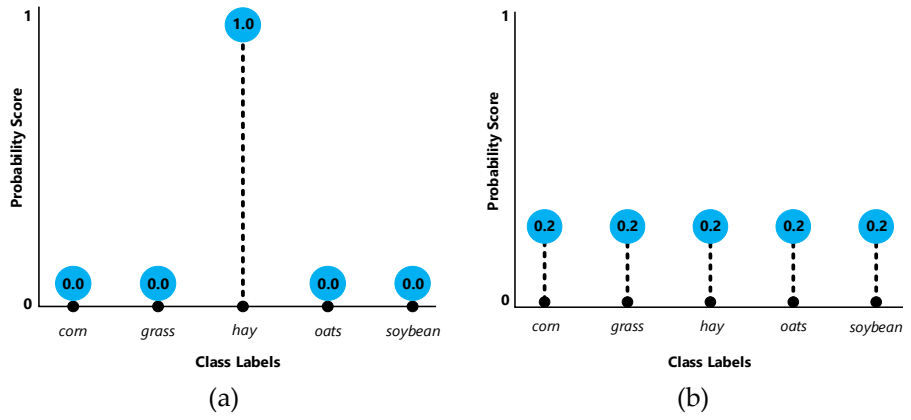


Figure 2.2. The best-case scenario for every pixel, (a) low uncertainty versus worst-case scenario, (b) high uncertainty. The other instances would be intermediate states of these two.

In this study, the uncertainty derived from deep learning neurons and popularity votes of random forest trees, were quantified using Shannon entropy (Shannon, 2001). Entropy summarizes the information from membership values of all classes using Eq. 2.1:

$$e_x = -\sum_{i=1}^{h} p_i \log_2 p_i \tag{2.1}$$

where p_i is the probability of class membership for *h* class labels. Further, the selection of the logarithm base is unimportant, as it only affects the units of entropy (Brown et al., 2009).

2.2.1.2. Deep neural network (DNN)

The deep learning algorithm applied in this research is based on R studio deep neural network (DNN) from H2O package (Candel et al., 2015), which is a feed-forward artificial neural network, trained with stochastic gradient descent using back-propagation. Here, multiple layers of hidden units were applied between the inputs and the outputs of the model (Hinton et al., 2012; LeCun et al., 2015; Li et al., 2013a).

Each hidden unit, *j*, typically uses the logistic function β (the closely related hyperbolic tangent is also often used and any function with a well-behaved derivative can be used) to map its output y_j using total input from x_j :

$$y_j = \beta(x_j) = \frac{1}{1+e^{-xj}} \tag{2.2}$$

For multiclass classification, such as our problem of hyperspectral image classification, output unit *j* converts its total input, x_j , into a class probability, p_j , by using a normalised exponential function named "*softmax*":

$$p_j = \frac{\exp(x_j)}{\sum_h \exp(x_h)} \tag{2.3}$$

where *h* is an index over all classes. DNNs is discriminatively trained by backpropagating derivatives of a cost function that measure the discrepancy between the target outputs and the actual outputs produced for each training case (Rumelhart et al., 1988). When using the softmax output function, the natural cost function *C* is the cross-entropy between the target probabilities *d* and the outputs of the softmax, *p:*

$$C = -\sum_i d_j \ln p_j \tag{2.4}$$

where the target probabilities, typically taking values of one or zero, are the supervised information provided to train the DNN algorithm.

2.2.1.3. Random forests as a benchmark

To measure and quantify DNN performance for uncertainty assessment of hyperspectral classification, we implemented the random forest (RF) algorithm applied to the same datasets (Liaw and Wiener, 2002). The RF algorithm provides an appropriate benchmark for assessing the performance of the DNN scheme because of its high performance found in hyperspectral data classification (Belgiu and Drăguţ, 2016; Chan and Paelinckx, 2008; Crawford et al., 2003; Ham et al., 2005; Lawrence et al., 2006). RF is also computationally efficient and suitable for training datasets with many variables and can solve multiclass classification problems (Mahapatra, 2014). We compared the uncertainty assessment results of DNN and RF using two different datasets.

2.2.1.4. RMSE of uncertainty assessment

Root mean square error (RMSE) is the standard deviation of the residuals (prediction errors). Here RMSE demonstrates standard deviation of prediction for correct and erroneous estimates of test dataset. In other words, it explains how concentrated the data is around the line of best fit considering entropy of correct and erroneous estimates:

$$RMSE = \sqrt{\sum_{i=1}^{n}(e-o)^2/n} \tag{2.5}$$

Where e represents estimated entropy value from '0' to '1' for minimum and maximum entropy after normalisation. o represents classification result for the observed values, which is '1' for erroneous predictions and '0' for correct answers. Here RMSE is applied as a goodness of fit for uncertainty assessment results. Therefore, the best-case scenarios would be those classification cases where the algorithm is at both maximum confidence and accuracy (e=0 and o=0) or minimum confidence and minimum accuracy (e=1 and o=1). The worst-case scenarios, however, occurs when the algorithm is at minimum confidence and maximum accuracy (e=1 and o=0) or vice versa (e=0 and o=1). Table 2.2 demonstrates the intuitions behind the proposed RMSE.

Table 2.2. The intuition behind the proposed RMSE

Best-case scenarios	***e***	***o***	**RMSE**	**Worst-case scenarios**	***e***	***o***	**RMSE**
Positive	0	0	0	Positive	0	1	1
Negative	1	1	0	Negative	1	0	1

* All other instances fall within intermediate states

2.2. Datasets

In this study, two widely used hyperspectral datasets including the Salinas (Kianisarkaleh and Ghassemian, 2016; Luo et al., 2017a; Yu et al., 2017) and Indian Pines (Li et al., 2017a; Yang et al., 2017; Yu et al., 2017) image datasets are used (Table 2.3) and divided into validation, train and test samples (Figure 2.3). Both datasets contain noisy bands due to dense water vapour, atmospheric effects, and sensor noise. These datasets are all available at http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

The Salinas image consists of 224 bands and each band contains 512 × 217 pixels covering 16 classes. It was recorded by AVIRIS sensor over Salinas Valley, CA, USA with a spatial resolution of 3.7 m, and the spectral information ranges from 0.4 to 2.5 μm. We used 204 bands, after removing bands of the water absorption features.

The Indian Pines dataset is also an AVIRIS image collected over the Indian Pines test site location, Indiana, USA. This dataset consists of 220 spectral bands in the same wavelength range as the Salinas dataset; however, four spectral bands are removed as they contain no data. This scene is a subset of a larger scene and it contains 145 × 145 pixels covering 16 ground truth classes. In the present research experiment, 20 spectral bands were removed because of the water absorption phenomena and noise.

Table 2.3. The major attributes of the hyperspectral datasets.

Dataset	**Sensor**	**Total bands**	**Excluded bands**	**Number of classes**	**Dimension**	**Resolution**
Salinas	AVIRIS	224	20	16	512 × 217	20 metre
Indian Pines	AVIRIS	224	24	16	145 × 145	20 metre

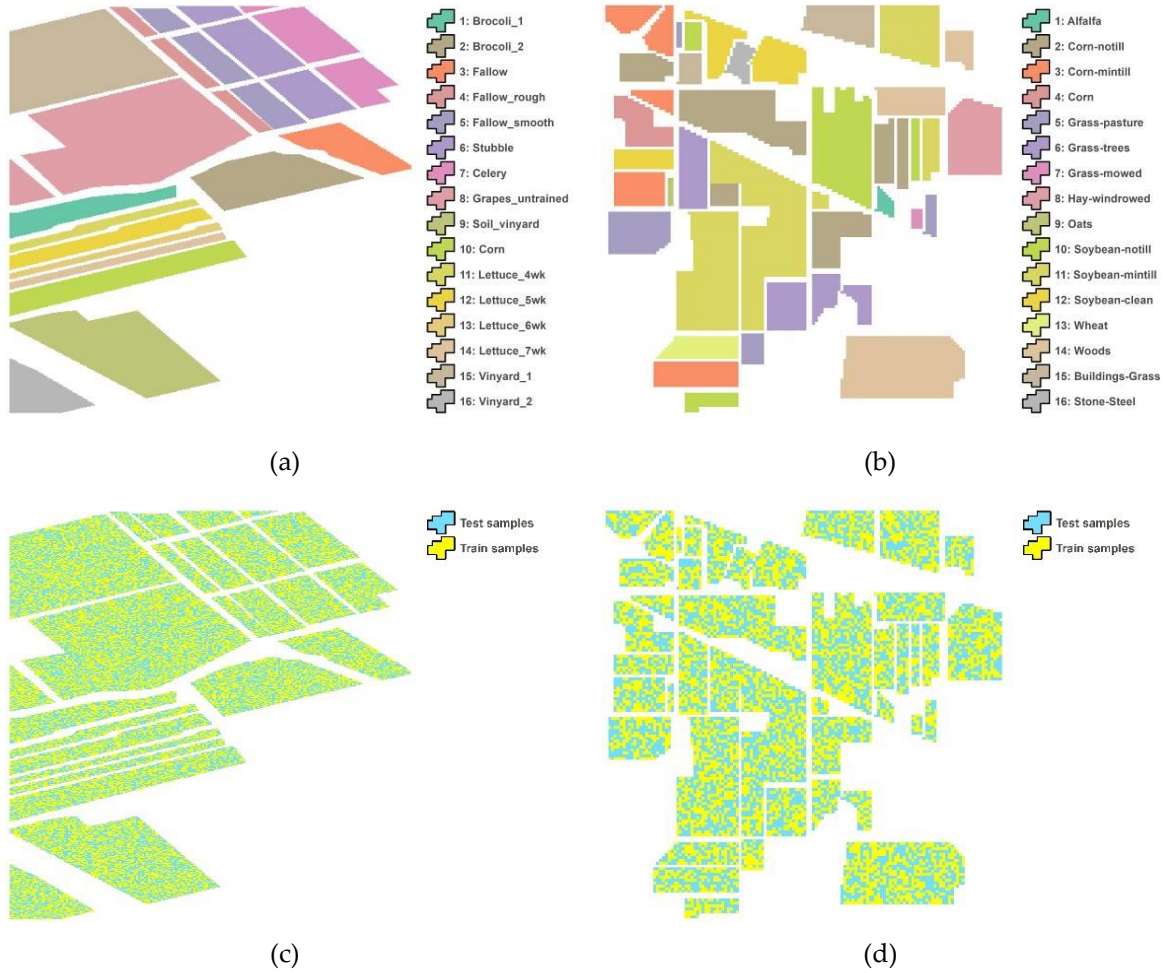


Figure 2.3. Ground truth data of two datasets including the Salinas (a) and the Indian Pines (b). The bottom images represent the location of the train and test data for the Salinas (c) and the Indian Pines (d).

2.3. Results

2.3.1. Salinas simulation experiments

The results of uncertainty assessment for Salinas dataset using DNN and RF are presented in Figure 2.4. However, to avoid redundancy in the representation of the results only half of the achieved uncertainty images are displayed (i.e. 10%, 30%, 50%, 70% and 90%). Regardless of the classification scheme and/or training sample size class 8 (8: Grapes_untrained) and 15 (15: Vinyard_1) belonged to the highest uncertainty level among all the available class labels. For both algorithms, this was followed by concentration of incorrect predictions within the high uncertainty areas, which are identified as false values in the mode of correct/incorrect classified test data based on all training samples from 10% to 100%.

RF and DNN were comparable in terms of achieved overall accuracy (OA) of classification for the majority of sample sizes while the areas covered with the high uncertainty values were less obvious within DNN results. This was observable for all corresponding sample sizes. Further, to quantify the capabilities of DNN and RF for uncertainty assessment and potential application as an estimate of accuracy, we then calculated the RMSE of every sample size (Eq. 2.5) applied for image classification for each algorithm. Following the mapping of uncertainty values, which represent uncertainty levels, we plotted the RMSE of the classification (y-axis) of the test data for various training sample sizes (x-axis). Here, lower RMSE values indicate better estimates of uncertainty (Table 2.2), and vice versa. For

the Salinas dataset, RMSE values for the DNN algorithm were lower than RF for all sample sizes while RMSE values derived from RF are more consistent (Figure 2.5).

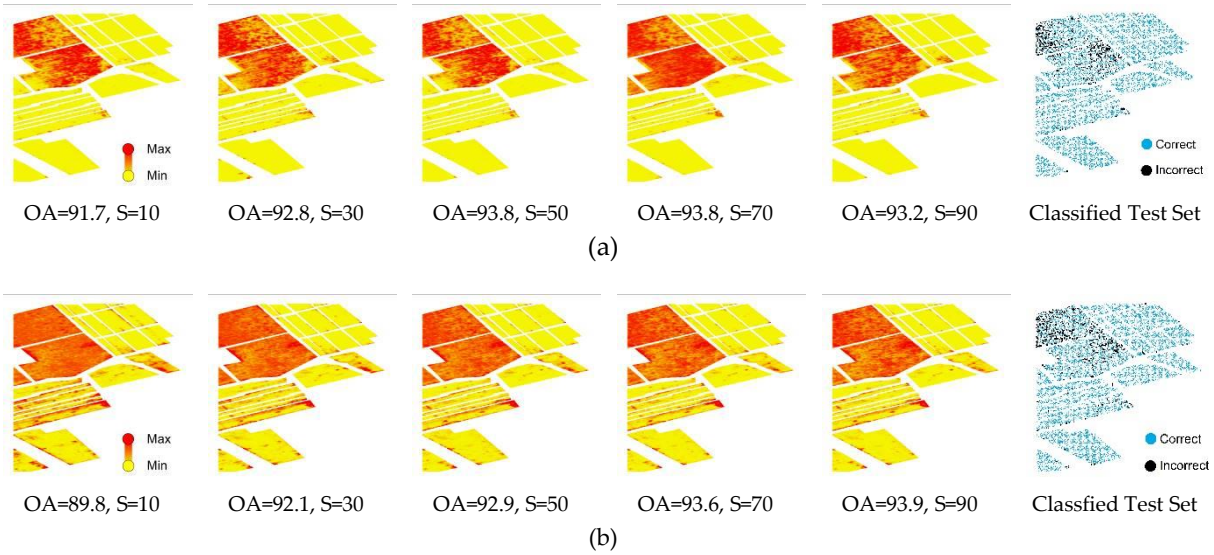


Figure 2.4. Results of uncertainty assessment for DNN (a) and RF (b) using different portions of training sample (S in %) and mode of correct/incorrect classified test data for the Salinas dataset. The estimated overall accuracy (OA in %) of the whole classification scheme is also demonstrated for selected portions of training sample.

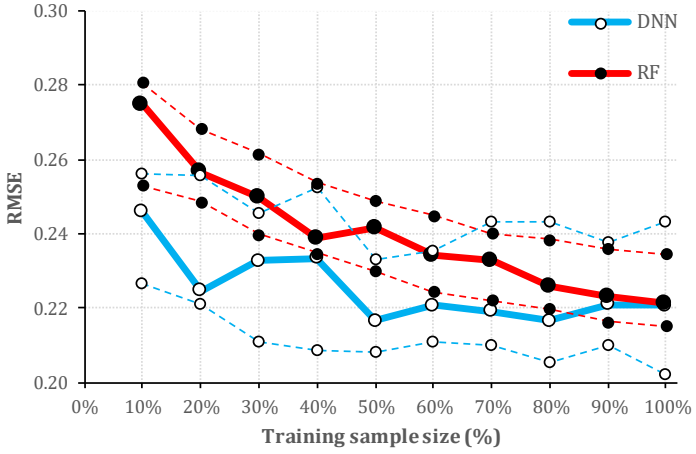
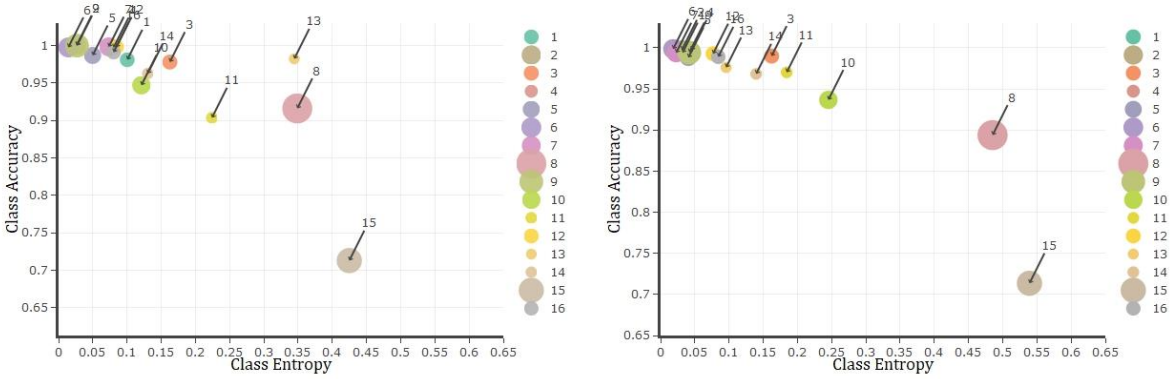


Figure 2.5. The estimated RMSE values of uncertainty assessment for test datasets (y-axis) where the algorithm is trained with different portions of the training sample (x-axis) of Salinas dataset. Dash lines represent minimum and maximum RMSE values for each sample sizes achieved in five consecutive simulation run.

Further, to better understand the capability of uncertainty measures as an estimate of accuracy, we plotted the correspondence between mean class uncertainty (i.e. entropy) and class accuracy (Figure 2.6). Nonetheless, to avoid unnecessary repetition of results, only the 50% training sample was plotted, which confirmed the accuracy of classification within the majority of image classes will be reduced by an increase in the uncertainty of pixels belonging to those classes and vice versa. In accordance with Figure 2.4, it was also demonstrated class 8 and 15 of Salinas dataset with highest mean uncertainty values belong to the least accurate estimation.



(a) (b)

Figure 2.6. Class entropy/uncertainty (*x*-axis) versus class accuracy (*y*-axis) plots of Salinas dataset using DNN (a, left) and RF (b, right) algorithms observed by applying 50% of training data. The bubble sizes represent the frequency of land-use class labels while bigger bubbles indicate the higher frequency and vice versa. Deep neural network 50% training sample size (a), Random forest 50% training sample size (b).

2.3.2. Indian Pines simulation experiments

The results of uncertainty assessment for Indian Pines dataset using DNN and RF were similar to the Salinas dataset. For both DNN and RF, classification uncertainty was reduced for larger training samples while the overall accuracy values of classification increased. However, this was less obvious for RF compared with DNN (Figure 2.7). In addition, the improvement of overall accuracy values with an increase in training sample size was more distinctive than the Salinas dataset. Remarkably, for every corresponding sample size, not only was DNN the more accurate algorithm but also displayed fewer pixels with high uncertainty values. The mode of correct/incorrect classified pixels demonstrated almost the same pattern for both algorithms while there were less misclassified pixels within the results of DNN algorithm.

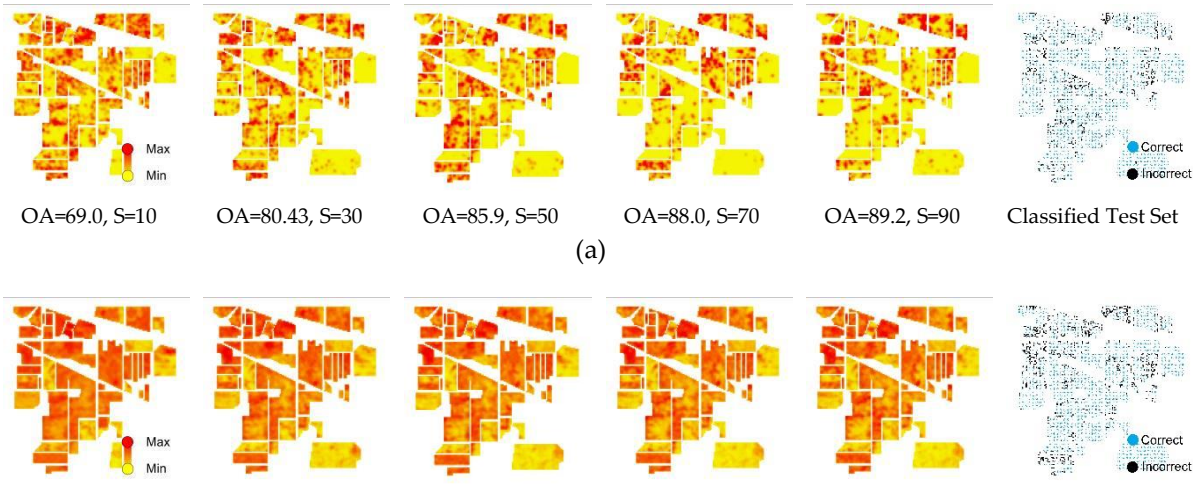


Figure 2.7. Results of uncertainty assessment for DNN (a) and RF (b) using different portions of training sample (S in %) and mode of correct/incorrect classified test data for Indian Pines dataset. The estimated overall accuracy (OA in %) of the whole classification scheme is also demonstrated for selected portions of training sample. Deep Neural networks (a), Random Forest (b).

The higher accuracy of DNN elevates the quality of implemented uncertainty assessment for locating correct/incorrect classifications for this dataset. Nonetheless, to quantify the difference in quality of uncertainty the assessment between the two algorithms, the RMSE values were estimated for every training sample. The RMSE values also confirmed the superiority of DNN for the majority of training sample sizes (Figure 2.8), in a way that less uncertainty was estimated for correct classified pixels while incorrect classified pixels were identified by more levels of uncertainty. However, the same as Salinas dataset, RMSE values derived from five consecutive simulation run of RF are more consistent.

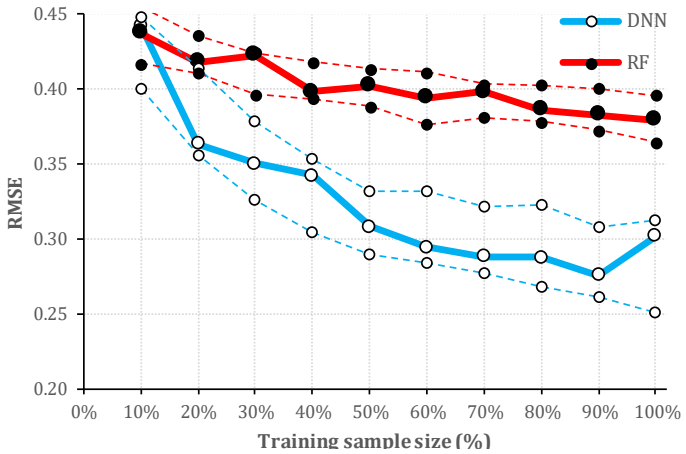
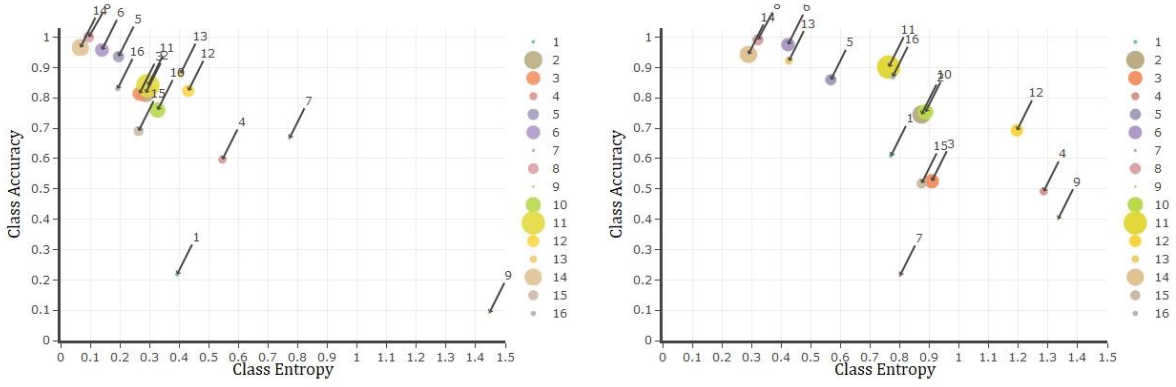


Figure 2.8. The estimated RMSE values of uncertainty assessment for test datasets (*y*-axis) where the algorithm is trained with different portions of training sample (*x*-axis) of Indian Pines dataset. Dash lines represent minimum and maximum RMSE values for each sample sizes achieved in five consecutive simulation run.

Finally, the correspondence between mean class uncertainty (i.e. entropy) and class accuracy of Indian pine dataset is demonstrated in Figure 2.9 for 50% training sample size using both DNN and RF algorithms. Similar to Salinas dataset results, the achieved results of Indian Pines demonstrated a negative relationship between uncertainty and accuracy for the majority of class labels.



(a) (b)

Figure 2.9. Class entropy/uncertainty (*x*-axis) versus class accuracy (*y*-axis) plots of Indian Pines dataset using DNN (a, left) and RF (b, right) algorithms observed by applying 50% of training sample size. The bubble sizes represent the frequency of land-use class labels while bigger bubbles indicate the higher frequency and vice versa.

2.4. Discussion

The evaluation of the two algorithms' performance as an estimate of accuracy was implemented based on uncertainty values of correctly and incorrectly classified pixels using Shannon entropy. Our results suggest the uncertainty of the probability values derived from neuron probabilities in DNN are preferred to those emanating from tree votes in RF for these two datasets. Considering the fact that

both of the algorithms were accurate algorithms, this is due to the promising capability of DNN to limit areas of high uncertainty within the vicinity of incorrectly classified pixels while RF failed to perform this way.

2.4.1. Comparing the quality of uncertainty assessment based on RMSE

With reference to the fact that both DNN and RF algorithms may achieve an overall accuracy above 70%, even for the minimum portion of training sample size (i.e. 10%), it was expected one algorithm may perform a better uncertainty assessment if it successfully limits the high uncertainty areas to the spatial vicinity of incorrectly classified pixels while highlighting the remaining areas as low uncertainty. This is regardless of achieved overall accuracy. Although the RMSE values derived from five consecutive runs of each algorithms indicate that results of uncertainty assessment using RF is more consistent compared with DNN. Nonetheless, comparing the results of uncertainty assessment, for both utilised datasets and every corresponding sample sizes demonstrates that areas of high uncertainty values were less abundant within the results of DNN algorithm compared with RF (Figure 2.4 and Figure 2.7). In other words, the uncertainty derived from DNN algorithm tended to be less than the uncertainty derived from popularity votes of random forest trees for the correct classified pixels. This may be due to the fact that DNN is optimized to reduce the difference between the predicted distribution and the true data generating distribution by minimizing the cross-entropy of the two probability distributions (Guo et al., 2017; Zhu et al., 2018). Therefore, the uncertainty assessment derived from DNN algorithm was a superior to RF combined with a better overall accuracy for these two datasets. This was validated by estimation of the modified RMSE values, using correct and erroneous estimates along with corresponding normalised uncertainty (Eq. 2.5), for both datasets and various training sample sizes (Figure 2.5 and Figure 2.8). However, more studies using different datasets are still required for generalizing the results.

2.4.2. Quality of uncertainty assessment for different sample sizes

For both algorithms and both datasets, larger training samples were found to be more beneficial for uncertainty assessment. The RMSE of uncertainty estimates, which was applied as a goodness of fit to assess the quality of uncertainty maps, decreased from the initial (10%) to final (100%) training sample sizes (Figure 2.5 and 2.8). However, this improvement was more obvious for DNN compared with RF. This may be due to a different formulation of RF and DNN algorithms, which is affecting the performance of the two algorithms for uncertainty assessment. Usually, training sample size has a crucial role in classification accuracy (Lu and Weng, 2007); thus it will also affect the uncertainty assessment process. The increased training sample size will typically increase the performance of an algorithm from random sampling (Foody and Mathur, 2006; Pal and Foody, 2010), but not all algorithms will be improved at the same level with a larger sample size.

Although RF can also benefit from a larger training sample by extracting more binary rules (Li et al., 2014), DNN may achieve a better performance. For DNN, the ratio of uncertainty assessment improvement followed by larger training sample size and more accurate classification depends on the abundance of contextual information per-pixel in the target dataset (Heydari and Mountrakis, 2018). In this regards, DNN benefits more from larger training sample ratio compared with the traditional algorithms, specifically using the rich spectral and contextual information in hyperspectral data for image classification (Chen et al., 2014). As many extensive experimental results confirm the excellent performance of the deep learning-based algorithms matched with rich spectral and contextual information (Zhang et al., 2016), our study suggests this is also beneficial to increase training sample to achieve a better uncertainty assessment result.

2.4.3. Uncertainty vs accuracy

The existing uncertainties at different stages of the classification procedure influence classification accuracy (Canters, 1997; Lu and Weng, 2007). Therefore, understanding the relationships between the

classification uncertainty and accuracy is the key successful contribution of an estimate of accuracy for image classification. Although low uncertainty classification instance is accompanied with high accuracy, some exceptions may apply to the high uncertainty, which usually belongs to low accuracy estimates. Thus, incorrect predicted class labels are usually located inside high uncertainty areas with very few exceptions within low uncertainty regions while correct classified pixel overlay the low uncertainty areas (Figure 2.4 and Figure 2.7). In this research, for both applied datasets, the existing correspondence between uncertainty and accuracy was better identified using DNN algorithm. Having said that, in our study, DNN is demonstrated more potential in uncertainty assessment for hyperspectral image classification. Following accurate classifications combined with minimising high uncertainty areas, DNN not only offers a lower rate of RMSE but also offers a higher contrast between low and high uncertainty areas.

At a wider scale, considering mean class uncertainty against the class accuracy of test data, it was revealed that usually a lower uncertainty value of a class is followed by a higher accuracy (Figure 2.6 and Figure 2.9). In other words, as low uncertainty indicates the probabilities of potential class labels for a pixel are not equal (i.e. unimodal distribution). This simply specifies that based on the available distribution of potential labels and their probability values (Figure 2.2), defined by either deep learning neurons or tree votes, usually one of the potential class labels (i.e. 16 labels for each applied datasets) has a significant preference to be selected as the estimated label. Accordingly, the concentration of low uncertainty values corresponding to every pixels of a desired class label is anticipated by an acceptable accuracy of classification. In terms of higher values of mean uncertainty for a class, the class accuracy will be reduced due to abundance of high uncertainty estimates within that class. Although, the imperfect correspondence between mean class uncertainty and class accuracy hinders us from developing a mathematical formulation to use uncertainty as an estimate of accuracy. Nonetheless, one may expect that image classification within a class with low mean uncertainty value will be more accurate compared with other classes with higher mean uncertainty.

2.5. Conclusion

Due to the weaknesses of the traditional approaches of map accuracy assessment based on a confusion matrix, many uncertainty assessment approaches are being developed as accuracy estimates. In terms of supervised methods, we compared deep neural network (DNN) with random forest (RF), where an estimate of accuracy is defined by the entropy of all potential probabilities/votes toward different class labels for a pixel, as an uncertainty measure. In this research, entropy was applied to encode the measure of uncertainty, which is applicable to any dataset including hyperspectral image datasets. Considering the results of uncertainty assessment, for both Salinas and Indian Pines datasets, DNN outperformed RF for the purpose of uncertainty assessment. However, the superiority of DNN algorithm was more obvious when applying Indian Pines dataset, as well as larger training sample sizes. This was due to, less-abundant high uncertainty values throughout the classified dataset compared with RF for every corresponding training sample size while having a comparable or better overall accuracy. Nonetheless, the achieved uncertainty maps of DNN can facilitate the application of hyperspectral image classification products by alerting map users about the spatial variation of classification uncertainty over the entire mapped region as an estimate of accuracy.

3. A robust rule-based ensemble framework using mean-shift segmentation for hyperspectral image classification

The current chapter aims to develop a robust rule-based ensemble framework using mean-shift segmentation for hyperspectral image classification. The proposed framework consists of multiple rule sets with rules constructed based on different class frequencies and sequences of occurrences where Shannon entropy is derived for assessing the uncertainty of every rule and the subsequent filtering of unreliable rules. This chapter, published in **Remote Sensing MDPI**, corresponds to objective number two and is grouped in modelling uncertainty Type A as explained and outlined in the thesis structure.

Shadman Roodposhti, M.; Lucieer, A.; Anees, A.; Bryan, B.A., 2019. A Robust Rule-Based Ensemble Framework Using Mean-Shift Segmentation for Hyperspectral Image Classification. Remote Sens. 11: 2057

Abstract— This paper assesses the performance of DoTRules—a dictionary of trusted rules—as a supervised rule-based ensemble framework based on the mean-shift segmentation for hyperspectral image classification. The proposed ensemble framework consists of multiple rule sets with rules constructed based on different class frequencies and sequences of occurrences. Shannon entropy was derived for assessing the uncertainty of every rule and the subsequent filtering of unreliable rules. DoTRules is not only a transparent approach for image classification but also a tool to map rule uncertainty, where rule uncertainty assessment can be applied as an estimate of classification accuracy prior to image classification. In this research, the proposed image classification framework is implemented using three world reference hyperspectral image datasets. We found that the overall accuracy of classification using the proposed ensemble framework was superior to state-of-the-art ensemble algorithms, as well as two non-ensemble algorithms, at multiple training sample sizes. We believe DoTRules can be applied more generally to the classification of discrete data such as hyperspectral satellite imagery products.

3.1. Introduction

Image classification is a vital tool for generating maps for environmental monitoring (Chan and Paelinckx, 2008). While for decades, multispectral imagery archives have been used to produce thematic maps, hyperspectral imagery is potentially a better option because of the higher spectral resolution. Hyperspectral images, which often contain more than 50 bands of continuous spectral information (Adep et al., 2016), can provide considerably more spatial and spectral information about the visible objects in their recorded field of view than multispectral imagery (van der Meer et al., 2012). Because of the quality of information, hyperspectral images are widely used in applications such as precision agriculture (Zarco-Tejada et al., 2016), biotechnology (Wakholi et al., 2017), mineral exploration (Rodger et al., 2012), and land-cover investigations (Chen et al., 2017). These various types of applications have generated interest in hyperspectral image classification that has grown rapidly during the past two decades, with significant progress (Ma et al., 2016).

Up to now, many popular machine-learning algorithms have been applied in hyperspectral image classification. These include instance-based (Huang et al., 2015), regression (Khodadadzadeh et al., 2014), regularization (Peng et al., 2016), decision tree (Goel et al., 2003), probabilistic (Shao et al., 2017), reinforcement learning (Zhong and Zhang, 2012), dimensionality reduction (Reshma et al., 2016), ensemble (Naidoo et al., 2012), Bayesian (Kayabol and Kutluk, 2016), maximum margin (Gao et al., 2015), evolutionary (Feng et al., 2016), clustering (Huang et al., 2015), association rule learning (Guo et al., 2006), artificial neural network (Awad, 2014; Goel et al., 2003; Yu et al., 2017) and deep learning (Li et al., 2017b) methods (see Figure 3.1). Regardless of the classification performance, many of these algorithms act as black-boxes, resulting in a poor recognition of the classification structure and robustness owing to the high-dimensionality of the data (Castelvecchi, 2016; Goodfellow et al., 2015).

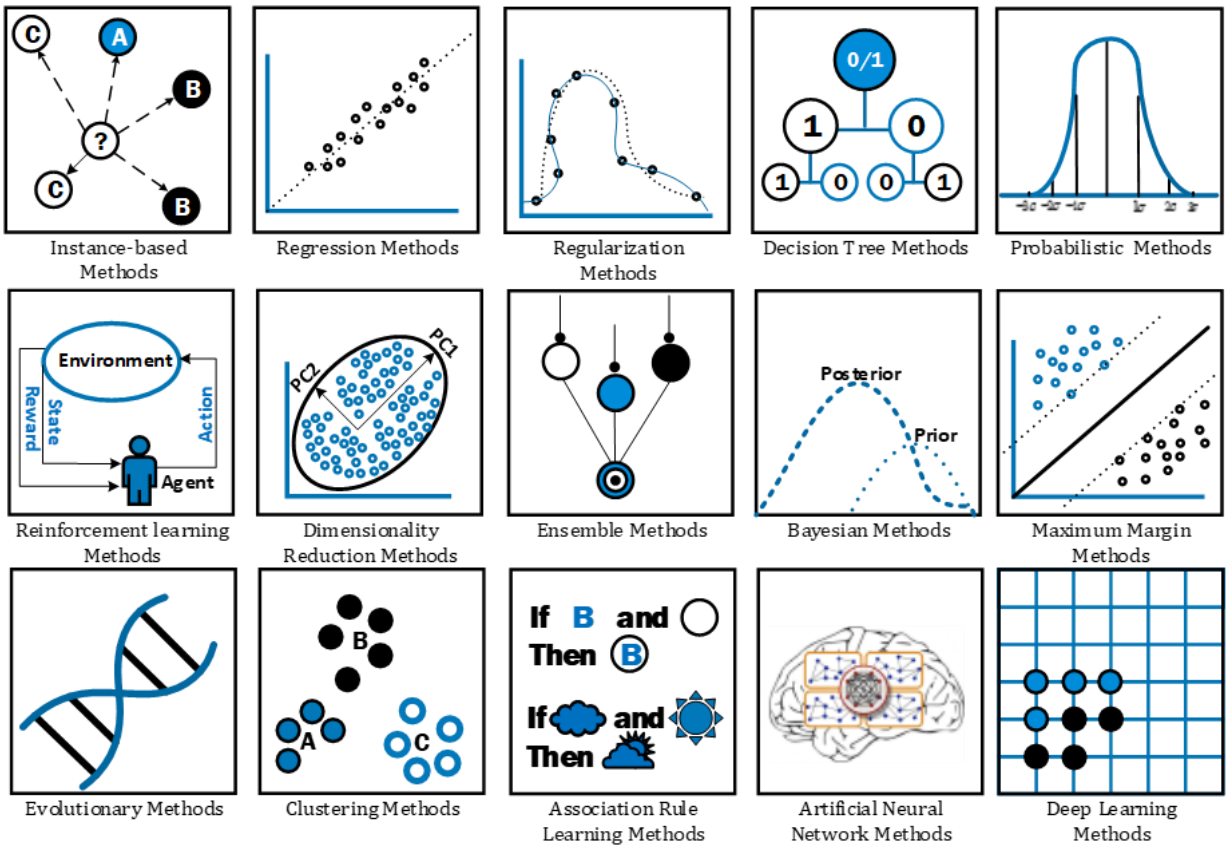


Fig. 3.1. Visual illustration of different categories of machine learning methods used for image classification.

Recently, ensemble classification methods have received more attention from the machine learning community, resulting in their increased popularity in different applications such as hyperspectral image classification (Ayerdi and Graña, 2016; Ayerdi et al., 2015; Uslu et al., 2017). Nonetheless, as opposed to other black-box classification algorithms, rule-based ensembles have demonstrated the ability to inform the interpretation of classification schemes (Tseng et al., 2008). Rules are very general structures that offer an easily understandable and transparent way to find the most reliable class allocation (Russell et al., 2003). The inferred logic of the model structure obtained by rule-based methods can be dissected, deciphered and applied out-of-the-box to new homogeneous classification problems. This is a major motivation, and it makes rule-based approaches more desirable compared with black-box approaches, even at the potential cost of a reduced classification accuracy (Bauer and Steinnocher, 2001). This paper presents a simplified and novel rule-based ensemble framework based on the mean-shift and uncertainty assessment as a hyperspectral image classification tool, and we compare its performance against other state-of-the-art ensemble algorithms, where the mean-shift application is exclusive to the proposed framework. For the sake of simplicity throughout the paper, the proposed framework is referred to as DoTRules (Dictionary of Trusted Rules).

Here, we present DoTRules for hyperspectral image classification to provide a better and more transparent understanding of classification schemes, as well as accurate and robust classification performance. This adds to the growing literature of ensemble methods applied to the classification of hyperspectral data (Benediktsson et al., 2008; Ceamanos et al., 2010; Xia et al., 2018), especially those aimed at improving the performance of classification with acceptable clarity (Rokach, 2010). DoTRules is based on rules and uncertainty assessment. It was first introduced and applied to the calibration of land-use/cover change simulation models (Shadman et al., 2017). We assess the performance of the DoTRules algorithm as a novel rule-based classification framework modified to employ a bagging approach in order to boost accuracy. This accuracy boost is implemented by applying a thresholding

assignment in order to extract trusted rules and then employing a novel voting approach to extract the class label recommended by the more trusted rules.

DoTRules extracts different subsets of training data from the full dataset, which can then be incorporated into boosting accuracy using a bagging approach designed to improve stability and accuracy. DoTRules has been found to perform well at modelling discrete data (Roodposhti et al., 2019a). Since satellite imagery products inherently contain discrete digital numbers (DNs), DoTRules can work natively with them, quantifying the likelihood of belonging to a certain map class. It identifies classification rules and quantifies their frequencies so that some will be more influential than others. It also handles null values, which originate from unmatched rules between training and test samples. In addition, the uncertainty of every recognised classification rule is quantified using Shannon entropy. In simple terms, it scrutinises the uncertainty of each classification rule prior to assigning class labels based on their uncertainty value, so that the overall accuracy of classification can be improved. This not only results in boosting accuracy but also enables data analysts to spatially map every unique rule's uncertainty. In terms of applying DoTRules, every pixel of the target hyperspectral dataset corresponds to one rule from each rule set, and, after quantifying uncertainty, only the most competitive one is selected among all of the corresponding rules for a target pixel. Thus, as opposed to many other methods, DoTRules is not a black-box method, as the attributes and characteristics of every single rule can be openly observed. In addition, by quantifying the uncertainty of every rule we can then anticipate their hit ratio. This provides a tool for the spatial segregation of more reliable/accurate classified boundaries from less reliable/accurate ones prior to image classification.

The main objectives of this study are to: (1) demonstrate DoTRules as an accurate and transparent rule-based ensemble framework for hyperspectral image classification; (2) map the uncertainty of every unique classification rule as an estimate of the rules' hit ratio. This highlights the contribution of this paper, i.e., developing an accurate and transparent rule-based ensemble algorithm that provides a prior estimate of classification accuracy at the pixel level. Mapping the spatial distribution of classification accuracies is considered extremely beneficial for enhancing the capabilities of a classifier used as a land-use and land-cover map production tool based on satellite imagery (Congalton, 1991; Khatami et al., 2017). Here, we describe the modified version of DoTRules for hyperspectral image classification, before demonstrating its application in three different study areas. We quantify the accuracy of DoTRules for hyperspectral image classification, and compare the results against some popular state-of-the-art ensemble approaches, i.e., extreme gradient boosting (XGBoost) (Georganos et al., 2018; Loggenberg et al., 2018), random forest (RF) (Chan and Paelinckx, 2008; Crawford et al., 2003; Gislason et al., 2006; Ham et al., 2005; Lawrence et al., 2006), rotation forests (RoFs) (Feng and Bao, 2017; Xia et al., 2015; Xia et al., 2014; Xia et al., 2016), regularised random forest (RRF) (Izquierdo-Verdiguier et al., 2017; Mureriwa et al., 2016), as well as two non-ensemble algorithms, namely, support vector machine (SVM) (Bazi and Melgani, 2006; Cui and Prasad, 2015; Fauvel et al., 2008; Lv et al., 2016; Tarabalka et al., 2010), and deep belief network (DBN) (Chen et al., 2015; Li et al., 2018a) as the classic deep learning method. Although SVM and DBN are not ensemble methods, they are included in our comparison because of their popularity, as they have been repeatedly used in recent hyperspectral image classification studies using Indian Pines, Salinas and Pavia University datasets. Finally, we discuss the advantages and disadvantages of the proposed approach for hyperspectral image classification.

3.2. Methods and datasets

3.2.1. DoTRules

DoTRules is based on a dictionary of trusted rules. It is designed for prediction when a large amount of discretised data are involved. However, it may also be applied to continuous data after discretisation. This is similar to the RF (Breiman, 1996; Breiman, 2001) method insofar as rule sets are used to select the mode response (i.e. most frequent class label). However, instead of generating random trees, the DoTRules operates by constructing many corresponding rules for every pixel (i.e. feature vector), which are derived from different rule sets. Designing a rule-based system is aimed to

clarify the ensemble scheme where every rule corresponds with one or more pixels. Each rule set is generated from a different combination of predictor variables in the ensemble run. For every unique rule, the most frequently occurring class label is assigned that carries the highest probability of occurrence. However, as there are many rule sets, there may be many matching rules with defined class labels for a single data sample. To get the best (i.e. final) class label, a weighted majority filter (weighted mode) is applied on every available corresponding rule for a single data sample after the elimination of unreliable rules. The weighted majority filter is urged to put more emphasis on those rules that are assembled by more components (i.e. matching variables) with less generalised class labels. The DoTRules procedure consists of the following steps implemented in R (R Core Team, 2017):

STEP 1: *Segmentation analysis*

As per requirement of proposed framework for systematic rule setting, a data segmentation or segmentation analysis should be applied to each predictor variable $J=\{j_1,j_2, \dots , j_n\}$ before classification, where J is a defined set of *spectral bands* / *band compositions*, but not necessarily every spectral bands or possible compositions of a hyperspectral dataset. Thus, the homogeneous digital numbers (DN) of the hyperspectral satellite image will be converted to segments. This is intended to partition m observations of original image into S segments for each j in J, in which each DN in each segment (ideally) share some common trait. Although various types of segmentation or even clustering algorithms can benefit the proposed classification framework, here, the "mean-shift" segmentation algorithm was applied (Cheng, 1995). The mean shift algorithm (Cheng, 1995; Fukunaga and Hostetler, 1975) is a recursive algorithm that allows us to execute a nonparametric mode-based segmentation. This is performed by data segmentation based on a kernel density estimate of the probability density function associated with the data-generating process. The core motivation factors for applying mean-shift algorithms is the fact that it is model-free and does not assume any prior shape on data segments. Furthermore, it is robust to outliers and does not require setting number of segments.

In its standard form, the mean-shift algorithm works as follows. A set of DN values from $x_1,\dots,x_m$ is observed for each spectral band $J=\{j_1,j_2, \dots , j_n\}$. A kernel function *ker f* is fixed and a bandwidth parameter σ and the update rule applied:

$$x \leftarrow \frac{\sum_{i=1}^{m} \ker f\left(\left\|\frac{x_i - x}{\sigma}\right\|\right) x_i}{\sum_{i=1}^{m} \ker f\left(\left\|\frac{x_i - x}{\sigma}\right\|\right)} \tag{3.1}$$

where σ is a bandwidth parameter. The fundamental parameter in mean-shift algorithms is the bandwidth σ, which determines the number of segments (Silverman, 2018). Furthermore, regions with less than some pixel-count C may be optionally eliminated. To account for different spatial and spectral variances it is practical to choose a kernel window of size $\sigma = \sigma_s, \sigma_r$ with differing radii σ_s in the spatial and σ_r in the range domain. The statistics literature has developed various ways to estimate the bandwidth. One of them is the adaptive mean-shift where you let the bandwidth parameter vary for each data point. Here, the σ parameter is calculated using kNN algorithm (Carreira-Perpinán, 2015). If $x_{i,S}$ is the k-nearest neighbour of x_i then the bandwidth is calculated as:

$$\sigma_i = \left\|x_i - x_{i,S}\right\| \tag{3.2}$$

Here, the aim of the segmentation analysis is to summarise input data and then minimise the required number of rules for correctly classifying pixels to their corresponding class label. As more accurate segments will improve the classification results, it is beneficial to apply the segmentation analysis on spectral band compositions composed of less similar spectral bands (e.g. within multidimensional space). Thus a pairwise dissimilarity measure $dis(j_i, j_n)$ between spectral bands j_i and j_n, for $1 \le i, j \le n$ (Huband et al., 2005) can be applied to achieve more robust segments.

STEP 2: *Formatting the data*

In order to avaoid mixining segments (S) values during concatenation phase for rule implementation in later steps, data segments should be formatted. Following data segmentation,

considering the maximum number of segments (*S*), the obtained data from step one should preferably be converted to two- (*S < 100*), three-digit (100 <= *S < 1000*) numbers or more. This is a requirement prior to the rule implementation. Hence, if a maximum value of *S* is less than 100, data should be formatted in a two-digit format (i.e. 3 = 03 or 26 = 26), while if maximum value of *S* is equal to or greater than 100 and less than 1000, then data should be in a three-digit format (i.e. 3 = 003 or 26 = 026), and so forth.

STEP 3: *Splitting data into training and test samples*

Both our training and test sets will be in tabular form, consisting of a set of pixels $I=\{i_1,i_2, .., i_m\}$. Each pixel i in I has a value x_{ij} for each predictor variable J. Simply, x_{ij} is the converted segment value of sample i in I for j from J. Thus, for each predictor variable j, x_{ij} can adopt one of a fixed set of possible values ≤ S. Each pixel i has a corresponding class label $l_i \in$ L=$\{l_1, l_2, \ldots, l_h\}$, which are also discrete semantic attributes from the global set of class labels such as corn, grass, oats, etc. It must be noted that to implement ensemble learners using DoTRules, it is needed to derive z sub-sets of the training dataset to construct different rule sets D containing individual classification rules d. This consists of all the available pixels in the primary training dataset but includes a different (random) combination of j in the feature vector.

STEP 4: *Creating a rule set*

For every z^{th} sub-set of the training set, we will concatenate values of a pixel x_{ij} for every j in J to form a rule set D. The concatenation of two or more characters is the string formed by them in a series (i.e. the concatenation of 001, 020, and 200 is 001020200). Equation 3.3 illustrates the pixel values for the cauterised predictor variables concatenated for each pixel (row) i, thereby creating a rule for each pixel in the corresponding subset of the training dataset.

$$D_z = \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{m1} \end{pmatrix} \| \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{m2} \end{pmatrix} \| \begin{matrix} \cdots \\ \vdots \\ \cdots \end{matrix} \| \begin{pmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{mn} \end{pmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{mn} \end{bmatrix} \tag{3.3}$$

Note that following the concatenation and extraction of rules (Equation 3.3), every rule within a specific rule set has maintained its single class label $l_i \in L$. We then aggregate duplicate rules where pixels have exactly the same values for all criteria, leaving an efficient new rule set of unique rules D'_z. The frequency of occurrence of all potential class labels $l_i \in L$ is then calculated for each unique rule d' in D'_z:

$$\begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_v \end{bmatrix} \rightarrow \begin{bmatrix} f(l_1) & f(l_2) & f(l_3) & \cdots & f(l_h) \\ f(l_1) & f(l_2) & f(l_2) & \cdots & f(l_h) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f(l_1) & f(l_2) & f(l_3) & \cdots & f(l_h) \end{bmatrix} \tag{3.4}$$

where v is the number of unique rules in D_z. The class label from set L with the highest frequency (i.e. the mode) is then assigned to each corresponding unique rule d'. The total number of rule sets D=[1, … , z] and the number of components in each rule set (i.e. the length of a rule) is user defined. Although the classification accuracy may increase by more rule sets, it will be at the expense of computation cost. In terms of rule length, the accuracy of classification may not increase necessarily by the implementation of longer rules, where longer rules —with more conditional components from J set— will model the training data too well (overfitting problem) resulting in less generalised responses for estimations of class labels, and vice versa (underfitting problem). In simple words, as the quantity of matching pixels in test dataset is inversely proportional to the length of rules, the longer rules with more components are more specific with fewer matches while shorter rules with fewer components are more general with many matches in the test dataset.

To ensure more accurate estimation, the default value of z is set 100 rule sets. Then, to avoid overfitting and underfitting issues, the number of predictor variables (j) used in every rule d within a specific rule set (length of rules within a considered rule set) is defined by a random function with a certain lower and upper bound defined by the user. This random function is called once, before creating every single rule set, to define the number and combination of components within that rule set. As the optimal combination of predictor variables is unknown, random band selection helps reduce the

potential for overfitting of the classifier. In this way, rules with various length will be implemented. The lower (*min*) and upper bound (*max*) for the length of rules (λ) in each rule sets D=[1, … , z] is a positive natural numbers defined by:

$$\lambda \begin{cases} \max(\lambda) \le n \\ \min(\lambda) > 0 \end{cases} \tag{3.5}$$

where n is the number of selected predictor variables in J set. The number of rule sets, min and max values of λ can be further optimised using a cross validation.

STEP 5: *Calculating and mapping rule entropy.*

The aim of this step is to assess the uncertainty value of each rule. In information theory, entropy is the quantitative measure of system disorder, instability and uncertainty, and may be used to forecast the trend of a specified system. Entropy indicates the expected amount of information contained (Shannon, 2001). Here, the entropy value of every unique rule d' from a rule set D_z' is calculated based on the frequencies of each possible class label (Equation 3.4) as:

$$e_{d'} = -\sum_{i=1}^{h} p_{l_i} \log_2 p_{l_i} \tag{3.6}$$

where $e_{d'}$ is the entropy of unique rule d' and p_{l_i} is the probability value of class label $l_i \in L$. Here h is the number of class labels in L. The general idea is that for a given rule, which may cover one or many pixels, the greater the probability of class membership for a given class label, the greater the certainty associated with that class. This provides a quantitative estimate of uncertainty for every single rule within different rule sets prior to assigning class labels. These estimates of uncertainty values can be applied to both the spatial mapping of rule uncertainty in classification, and to eliminating those unreliable rules with a high entropy value from different subsets and/or rule sets before combining votes. The spatial distribution of uncertainty is quantified by mapping the entropy of each unique rule back to the corresponding pixels. These estimates of uncertainties are extremely beneficial and can be considered even prior to assigning class labels to pixels. Every time that DoTRules is applied to a training data subset, a class label of the highest frequency is allocated, and the entropy of that rule is calculated.

STEP 6: *Eliminating unreliable rules within all rule sets.*

After assessing uncertainty value of each individual rule, unreliable rules (i.e. rules with high entropy) should be eliminated to improve the quality of voting outcome which directly affect the classification accuracy. Thus, every such rule d' (in D'_z), for which the $e_{d'}$ is greater than the user-defined threshold, is eliminated. In our study, we specified that if a rule has the $e_{d'}$ value greater than 0.3 and its corresponding pixels frequency less than α (to avoid randomness), then it is considered to be unreliable and eliminated accordingly. The α is can be calculated as follows, keeping the random chance for resultant entropy value less than 0.05%:

$$\alpha = Ceil\left(\frac{\ln(0.05)}{\ln(1/h)}\right) \quad for \quad h > 1 \tag{3.7}$$

where h is the number of class labels.

STEP 7: *Classify test dataset according to the DoTRules.*

Above, we described the process of creating the DoTRules and allocating the most likely class label for each rule based on frequency. In the same way, class labels can now be assigned for the study area using another subset of the primary training dataset (i.e. implementing more rule sets). Every time a new rule set is implemented, the same procedure is followed to establish rules for the test dataset. Each test data rules is then matched with their equivalent training rules in the DoTRules using a many-to-one matching algorithm and allocate the most likely class label to each test data rule. This will be repeated every time that a weak learner is being implemented from every single rule set.

STEP 8: *Handling null values*

There is always a possibility of encountering null records in the test dataset while using DoTRules. In such situations, new pixels in the test dataset present combinations of states for criteria not

encountered in the training data, which may increase *out of bag* error. Handling null values is inevitable to maintain the classification accuracy where in the proposed ensemble framework using mean-shift it is a combined procedure. First, all rules are sorted based on their similarity, then every single null value is assigned to the class label of its closest (i.e. most similar) rule, based on alphanumeric similarity of the constructed rules. However, the influence of these rules in combining votes is minimised as they are characterised by null entropy values.

STEP 9: *Combining votes*

To fulfil the classification procedure, this step assigns a final label to each pixel. To combine votes of each set of learners, all unreliable rules (with low or null entropy records) within every rule set must be removed using a thresholding approach. Afterwards, a mode filter is applied to the resultant class labels coming from sets of corresponding rules for each pixel. This mode function not only considers the frequency of class labels, but also considers the length of a rule as a weighted function. As a rule is formed by concatenating n number of predictor variables (j), therefore a rule that contains more predictor variables as components has a higher weight in the mode function. Nonetheless, if none of the recognised reliable rules, for a certain pixel in the test dataset, is matched by any corresponding rule from the various training rule sets (derived from subsets of training data), then the mode function will be applied to the corresponding labels of unreliable rules explained in STEP 8 with the same mode function.

STEP 10: *Calculating and mapping the hit ratio*

Calculating and mapping the hit ratio helps to visualise the spatial distribution of classification error. Similar to the entropy value, which is calculated for every unique rule based on the frequencies of each possible class label, we map the hit ratio of every unique rule in our combined results back to the original pixels. DoTRules is rule-based where every unique rule d' from a rule set D'_z corresponds to one or many pixels; thus, it is possible to calculate the classification hit ratio of those rules using Equation 3.8:

$$A_{d'} = \sum_{i=1}^{h} l_i^{+} \Big/ \sum_{i=1}^{h} l_i \tag{3.8}$$

Here, l_i^{+} is the sum of correct classified labels.

3.2.2. Rule uncertainty threshold

In using DoTRules for classification of hyperspectral imagery, the class label of a rule is also described by both its entropy value and the frequency of all potential class labels (Fig. 3.). Therefore, a rule can be considered reliable if its entropy is less than 0.3 bits, which is calculated at least for n potential class labels (frequency > α). However, it is important to note that among the reliable rules coming from the various rule sets for a certain pixel, those with a longer concatenated string (rule) will have more impact in combining final votes. This is mostly due to the fact that they are composed of more variables while they meet the same uncertainty threshold, hence can make more robust predictions. In other words, longer rules have fewer pixels with a specific class label while shorter rules have more pixels belonging to multiple class labels. The fewer the pixels shared between different rules are, the more accurate the classification results will be. As the estimated entropy values for distribution of response variables (class labels) with low frequencies are less reliable (Fig. 3.d) and may be due to random chance, a second threshold is applied to the frequencies of potential labels. This will further improve the quality of rule elimination process.

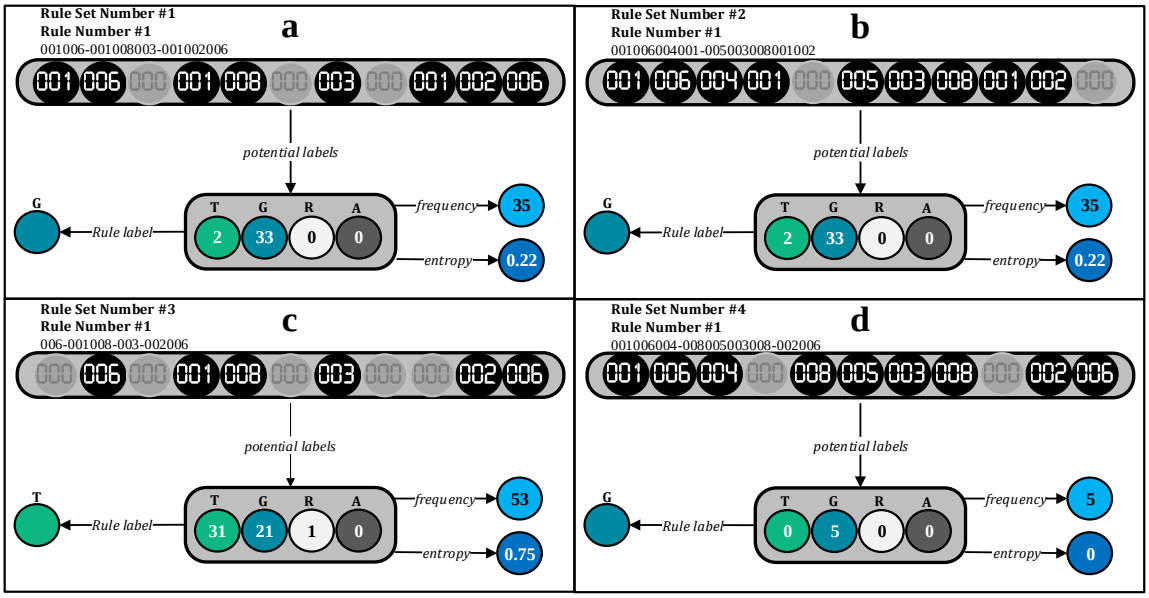


Fig. 3.2. Schematic demonstration of reliable (a and b) and unreliable (c and d) rules extracted using DoTRules. Black circles represent segment values of randomly selected spectral bands composing different rules for one target pixel. Considering rule sets number #1 and #2, the latter will have more impact in combining votes due to its larger length.

3.2.3. Comparing DoTRules with other methods

To measure and quantify DoTRules' performance, we implemented different classification algorithms, including XGBoost, RF, RoF, RRF, SVM, and DBN on the same datasets. These six algorithms are among the most popular methods for hyperspectral image classification, and they belong to three different categories of machine-learning methods. The first four algorithms are state-of-the-art ensemble methods, while SVM is a maximum margin classifier and DBN is a deep learning method. Thus, these methods provide appropriate benchmarks for assessing the performance of the DoTRules. XGBoost is an algorithm that has recently been dominating applied machine learning (Chen and Guestrin, 2016), and RF, RoF and RRF were selected because of both their natural similarity to DoTRules and performance in hyperspectral data classification (Chan and Paelinckx, 2008; Crawford et al., 2003; Gislason et al., 2006; Ham et al., 2005; Lawrence et al., 2006). They are also computationally efficient and suitable for large training datasets with many variables and can solve multiclass classification problems (Mahapatra, 2014). Furthermore, SVM (Gao et al., 2015; Golipour et al., 2016; Huang et al., 2015) and DBN (Chen et al., 2015; Li et al., 2018a) algorithms have demonstrated promising results in previous studies. We compared the overall accuracy (OA) and kappa coefficient (k) of DoTRules with these various algorithms for hyperspectral image classification, using three different datasets from Indian Pines, Salinas and Pavia University (Figure 3.3).

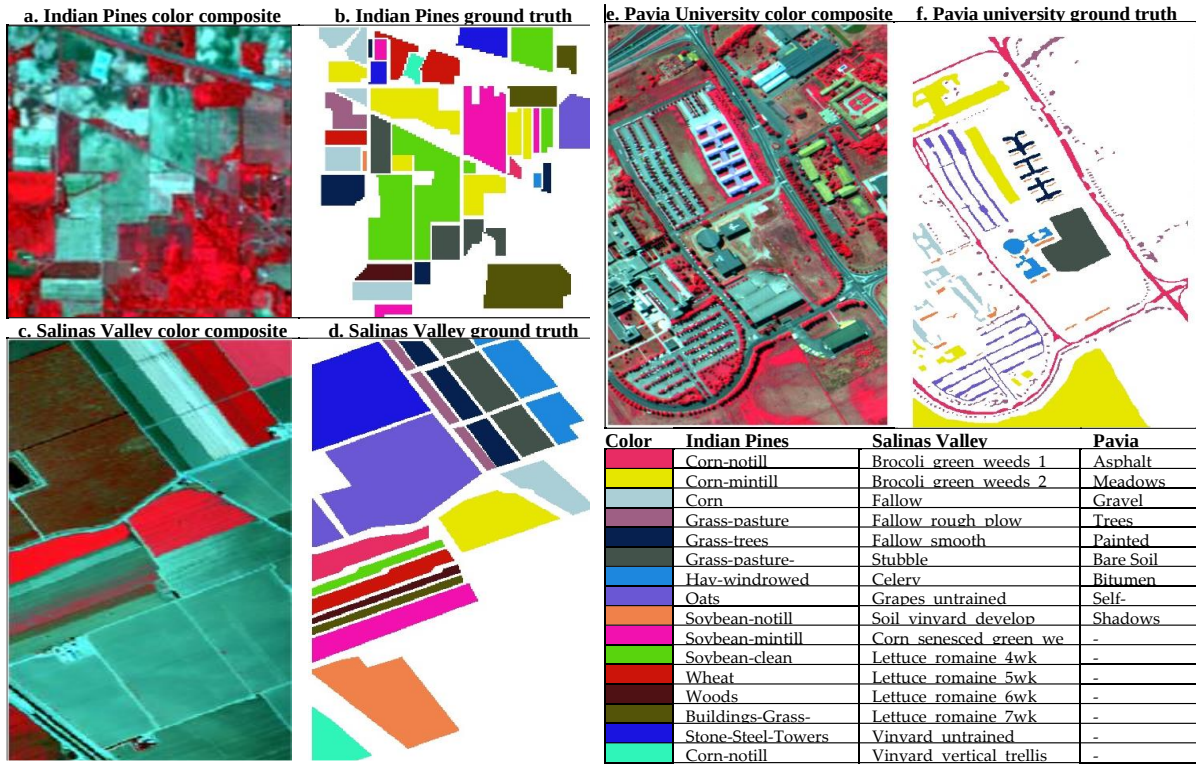


Fig. 3.3. False colour composites and ground truth images of the applied datasets for the image classification purpose including a, b) Indian Pines, c, d) Salinas and e, f) Pavia University dataset.

After tuning the required parameters of the above algorithms using the CARET package in R (Kuhn, 2008), a training process was implemented. In order to make a valid comparison, not only applicable to different study areas but also robust to variations of the portion of training and test sample sizes, different sample sizes of 1%, 5% and 10% were used. In addition, the overall accuracy value was taken as an average of five consecutive runs of each combination of algorithm and sample size. This was to avoid a sudden change in the overall accuracy value arising from changes in the training sample.

3.2.4. Datasets

DoTRules is tested using three hyperspectral image datasets (Fig. 3.) namely, Indian Pines (Yang et al., 2017; Yu et al., 2017), Salinas (Kianisarkaleh and Ghassemian, 2016; Luo et al., 2017a) and Pavia University dataset (Golipour et al., 2016; Yu et al., 2017). Both Indian Pinea and Salinas datasets contain noisy bands due to water vapour, atmospheric effects, and sensor noise. All the three datasets are available at http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. The mean spectral signatures of the three datasets is also demonstrated in Fig. 3..

The Indian Pines dataset is also an AVIRIS image collected over the Indian Pines test site location, Indiana, USA. This dataset consists of 220 spectral bands in the same wavelength range as the Salinas dataset; however, four spectral bands are removed as they contain no data. This scene is a subset of a larger scene and it contains 145 × 145 pixels covering 16 ground truth classes. In the present research experiment, 20 spectral bands were removed because of the water absorption and noise.

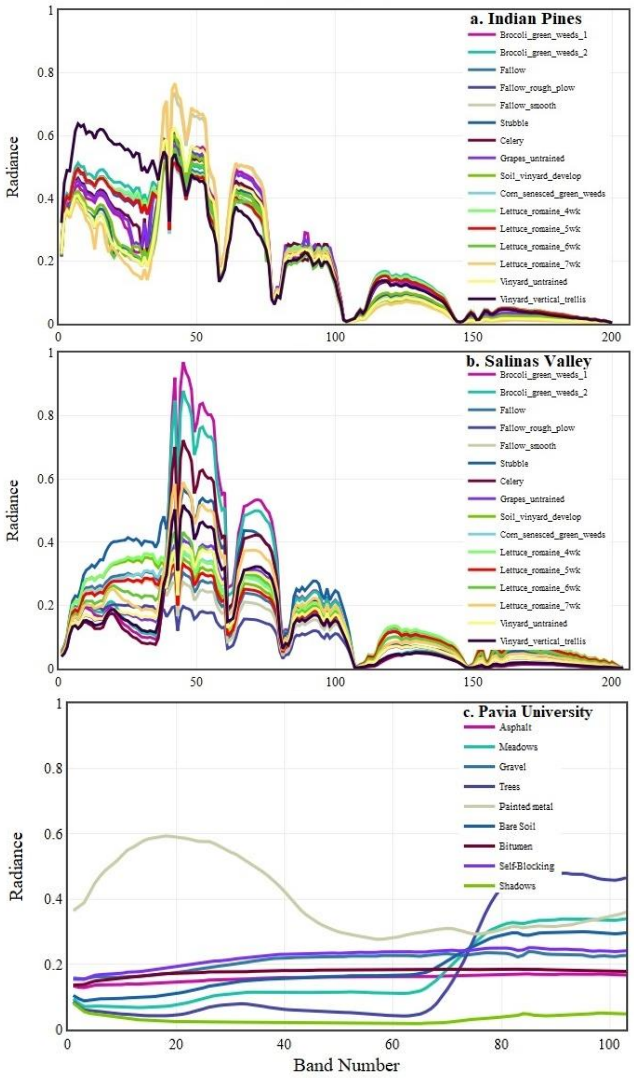


Fig. 3.4. Mean spectral signatures of a) Indian Pines, b) Salinas Valley and c) Pavia University datasets.

The Salinas image consists of 224 bands and each band contains 512 × 217 pixels covering 16 classes. It was recorded by AVIRIS sensor over Salinas Valley, CA, USA with a spatial resolution of 3.7 m, and the spectral information ranges from 0.4 to 2.5 μm. We used 204 bands, after removing bands of the water absorption features.

Pavia University dataset is collected by the Reflective Optics System Imaging Spectrometer (ROSIS) system that is compact airborne imaging spectrometer. It consists of 103 spectral bands after removing the noisy bands, and 610 × 340 pixels for each band with a pixel resolution of 1.3 m. The ground truth image consists of nine classes.

3.3. Results

3.3.1. Simulation experiments

For all three hyperspectral datasets, DoTRules was superior to all other algorithms in terms of the overall accuracy and kappa coefficient. However, considering the very low sample size (i.e., 1%) of the small-sized datasets (i.e., Indian Pines and Pavia University) it was not the most accurate approach. This is confirmed by the results of the accuracy assessment for the different sample sizes, which are averaged from five consecutive runs for a target sample size (Table 3.1).

Table 3.1. Accuracy assessment results of three applied datasets including the overall accuracy (OA%) and kappa coefficient (κ) for all applied methods including support vector machine (SVM), deep belief network (DBN), extreme gradient boosting (XGBoost), random forest (RF), rotation forests (RoFs), regularised random forest (RRF), as well as Dictionary of Trusted Rules (DoTRules). The maximum values are highlighted in bold.

	Train	Test	SVM	DBN	XGboost	RF	RoF	RRF	DoTRules
Indian Pines	1%	50%	62.2	56.0	52.9	64.8	**70.5**	58.8	68.6
			0.558	0.486	0.453	0.593	**0.650**	0.521	0.640
	5%	50%	75.0	73.0	69.8	69.3	77.9	64.6	**87.3**
			0.708	0.689	0.656	0.644	0.725	0.588	**0.855**
	10%	50%	81.0	78.6	75.0	73.4	84.9	72.3	**93.2**
			0.781	0.755	0.710	0.693	0.788	0.675	**0.928**
Salinas	1%	50%	90.6	87.7	89.0	86.6	89.9	88.1	**91.5**
			0.895	0.862	0.877	0.850	0.881	0.867	**0.906**
	5%	50%	92.3	92.2	90.8	90.3	91.9	90.1	**97.2**
			0.914	0.913	0.898	0.892	0.908	0.888	**0.969**
	10%	50%	93.3	92.3	92.1	91.5	92.9	90.6	**98.7**
			0.925	0.914	0.912	0.905	0.918	0.895	**0.986**
Pavia	1%	50%	**92.0**	86.7	81.6	81.8	84.9	81.6	79.1
			0.893	0.820	0.748	0.749	0.790	0.732	0.720
	5%	50%	93.0	93.0	88.7	87.6	88.2	87.3	**93.1**
			0.907	0.906	0.849	0.833	0.871	0.817	**0.909**
	10%	50%	94.4	94.2	91.2	89.4	91.4	88.9	**96.2**
			0.925	0.920	0.882	0.857	0.895	0.850	**0.951**

The classification results also demonstrates that DoTRules classification was able to closely match the spatial pattern of the ground truth image (Fig. 3.). These results were consistent across all three hyperspectral datasets. DoTRules was not only an accurate but also a transparent rule-based approach where the reliability (based on uncertainty) of each rule can be mapped. This is a desirable feature in remote sensing applications where the visual investigation of classification rules is informative.

3.3.2. Uncertainty mapping

As DoTRules is rule-based, and each unique rule with its specific entropy value corresponds to one or more pixels, it is possible to estimate and map the uncertainty of every unique rule back to those pixels. This is a preliminary product of DoTRules, before assigning a class label to every pixel.

To illustrate the applicability of the entropy map to locate areas belonging to a low versus high classification accuracy, entropy values above and below the applied threshold (e_d = 0.3, Equation 3.7) were mapped to segregate regions which have more reliable and less reliable classification responses (Figure 3.5). In this way, the DoTRules spatial uncertainty map can facilitate a better understanding of uncertainty in the classified product and the segregation of more and less reliable geographic areas before assigning class labels to every pixel of the test sample dataset. This provides clear spatial insight into the uncertainty of the classification at an early stage of the classification process.

In developing and applying DoTRules, we have found that a larger sample size offers a higher classification accuracy where the number of less reliable rules with higher levels of uncertainty is reduced. Conversely, a smaller sample size, with fewer rules detected in our rule sets, was less able to capture the complexity of the hyperspectral image classification. This is mainly due to the fact that for DoTRules, training samples should be enough to cover all possible forms of rules. Figure 3.5 demonstrates the rule uncertainty for the Indian Pines, Salinas and Pavia University datasets using 1%, 5% and 10% training sample sizes.

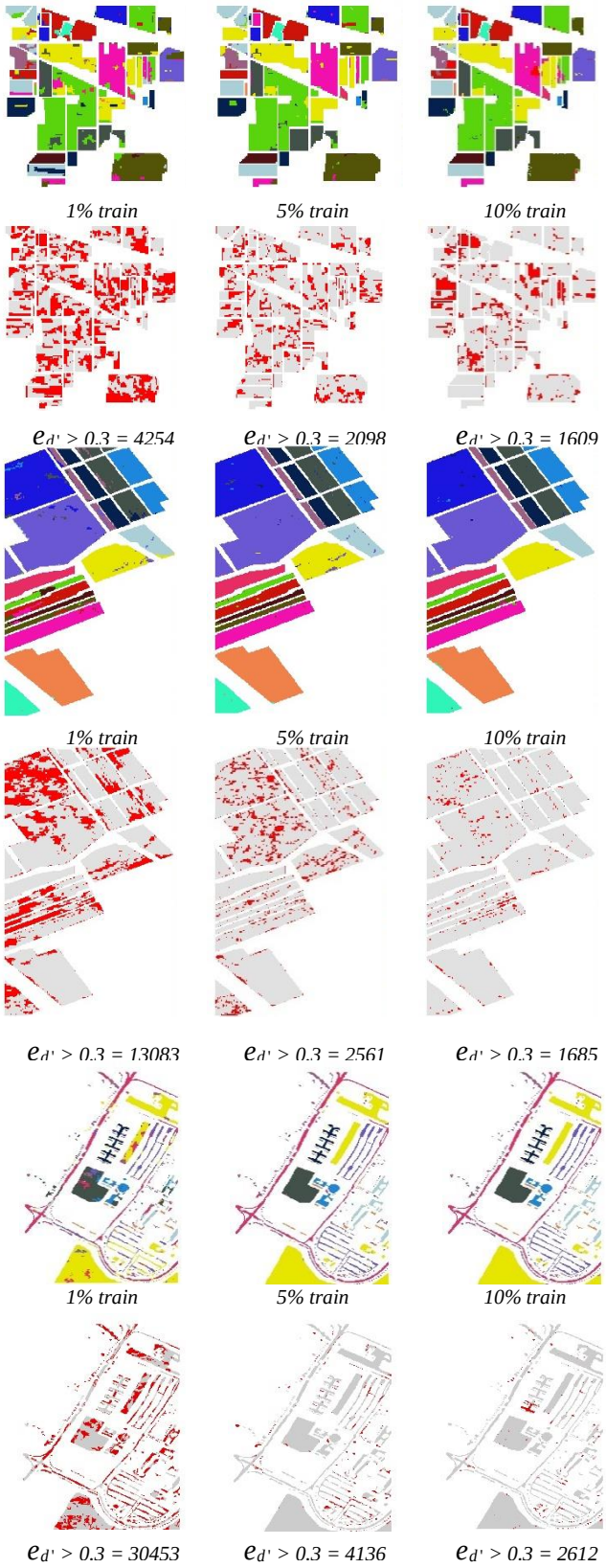


Fig. 3.5. DoTRules classification results and estimated pixel based $e_{d'}$, for Indian Pines, Salinas and Pavia datasets. Red pixels show the location of unreliable rules according to entropy thresholding ($e_{d'} > 0.3$, for α=0.05), while grey pixels are reliable rules above the threshold. Th red pixels are counted for each sample size.

3.3.3. Correspondence between uncertainty and hit ratio of rules

In general, where there is low entropy (i.e. low uncertainty) for a rule within our rule set, a more accurate classification rule may be achieved, as it has lower uncertainty and complexity. Simply, lower entropy means there is just one clear answer (the mode class label) for a rule while high entropy indicates a more uniform distribution of map class frequencies for that rule, which indicates less reliable classification rules with more complicated answers. Plotting estimated hit ratio values against entropy values of every constructed rule among our various rule sets demonstrate that the hit ratio of rules can be defined by a polynomial function of their entropy value (Fig. 3.), which is followed by a strong coefficient of determination for all three applied datasets.

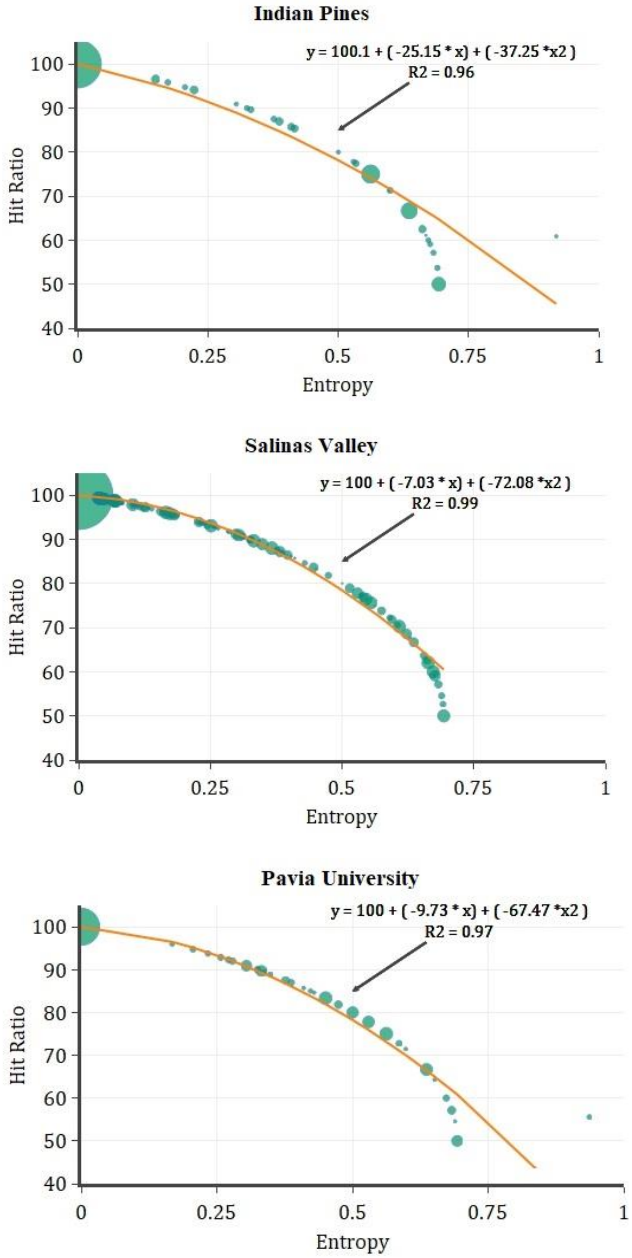


Fig. 3.6. The entropy versus the hit ratio of rules using DoTRules for a) Indian Pines, b) Salinas Valley and c) Pavia University dataset 10% training sample size. Bubble sizes show frequency of each rule among all corresponding rules from different rule sets before combining votes.

To further demonstrate the applicability of DoTRules' uncertainty product for the anticipation of rule exclusive hit ratio we then applied the derived functions based on the correspondence of hit ratio

and entropy of the training data to prediction the hit ratio of rules within the test datasets. The achieved RMSE values of predicted hit ratios based on entropy polynomial function is less than "one" (RMSE<1) for all applied datasets (Table 3.). Table 3. demonstrates that uncertainty product of DoTRules may be applied to estimate hit ratio of classification rules applied for the context of hyperspectral image classification with a satisfactory level of accuracy.

Table 3.2. Prediction of rules' hit ratio based on the corresponding entropy values for 10% training sample size

Dataset	R	R-squared	p-value	Train RMSE	Test RMSE
Indian Pines	0.978	0.958	2.20E-16	0.3261	0.0972
Salinas Valley	0.996	0.993	2.20E-16	0.0195	0.1087
Pavia University	0.985	0.971	2.20E-16	0.0142	0.0628

3.4. Discussion

In this paper, we have presented a rule-based ensemble framework based on a mean-shift segmentation and uncertainty analysis, referred to as DoTRules (a Dictionary of Trusted Rules), for hyperspectral image classification. DoTRules constructs many rule sets composed of corresponding rules for each pixel in a hyperspectral image to predict the class of the test samples. When applied to different datasets and sample sizes, DoTRules proved to be an effective strategy for the classification of hyperspectral imagery, with promising results compared to other established algorithms. Furthermore, DoTRules enables both rule uncertainty and hit ratio mapping, which is an advantage for the users of classified land-use and land-cover maps created from remote sensing imagery. Below, we discuss improvements in hyperspectral image classification achieved using DoTRules..

3.4.1. The overall accuracy of classification

According to our results, for all three applied hyperspectral datasets, the DoTRules ensemble framework was more accurate than the other applied classification algorithms for most training sample sizes (Table 3.1). This is due to the robust rule detection framework using mean-shift segmentation, where Shannon entropy is used to assess the uncertainty of individual rules for classification purpose. Here, the segmentation is done in a way where each DN in each segment (ideally) shares some common trait. This bears similarities with an object-oriented classification that involves the categorization of pixels based on the spatial relationship with the surrounding pixels. While pixel-based classification is exclusively based on the information in each pixel, object-based classification is based on information from a set of similar pixels (i.e., objects or image objects). Image objects are groups of pixels that are similar to one another based on the spectral properties (i.e., colour), size, shape, and texture, as well as context from a neighbourhood surrounding the pixels, in an attempt to mimic the type of analysis done by humans during visual interpretation. In addition, passing segment information to pixels and extracting reliable rules (i.e., low uncertainty rules) using minimum entropy through a voting system further preserves the high classification accuracy, especially when a representative training sample size is applied.

The observed increase in the overall accuracy of DoTRules' estimates when applying larger sample sizes may be due to an extra number of rules being detected and relatively fewer *null* records. Rules are very general structures that offer an easily understandable and transparent way to find the most reliable class allocation for a pixel (Russell et al., 2003). As opposed to decision trees, every rule corresponds to only one pixel. This is unique to DoTRules and a common criticism of XGboost, RF, RoF, RRF and similar black-box algorithms (Gislason et al., 2004; Palczewska et al., 2013). Users can access all rules and their corresponding information, such as the rule ID, components of a rule (segment class for every

selected band), true class label, probability (relative frequency) of every potential class label, rule entropy and hit ratio (accuracy) (Figure 3.2), while they are always connected to their corresponding pixels. This beneficial trait is highly valued in geoscience and remote sensing applications, especially in the context of land-use and land-cover mapping applications (Khatami et al., 2017; Lucas et al., 2007; Yang et al., 2015). To be able to assign every pixel to a map class, each pixel should have at least one matching rule from various rule sets. Logically, the number of recognised rules within each individual rule set will be increased by a consequent increase in the training sample size (i.e., 1% to 10%), while the number of null records derived from unmatched rules between the test and training dataset will be reduced. Therefore, the greater the number of trusted rules, the greater the capability of our proposed framework to allocate test pixels into their true class labels.

3.4.2. Quantifying and mapping the uncertainty of rules

While a few studies have successfully mapped the uncertainty of classification before image classification (Bryan et al., 2009; Khatami et al., 2017), one strength of DoTRules in hyperspectral image classification is its demonstrated ability to quantify the uncertainty of every identified transition rule using entropy values prior to the final classification (Figure 3.5). In other words, DoTRules was able to report the uncertainty of rules based on Shannon entropy, independent from the test dataset. The results from different hyperspectral datasets show that the lower the entropy value, the higher the hit ratio (Figure 3.6). Thus, considering the strong relationship between the entropy and hit ratio, it is possible to apply the entropy values as estimates of the hit ratio. The estimation of the rule uncertainty prior to the classification of a hyperspectral dataset aids in understanding the specific strengths and weaknesses of a classifier dealing with pixels containing a range of spectral information.

3.4.3. Quantifying hit ratio of rules

DoTRules demonstrated the ability to quantify the rule-exclusive hit ratio using their corresponding entropy values (Table 3.2). Thus, the uncertainty product based on the entropy values can be applied to segregate areas of less and more reliable prediction independently of the test data availability. Thus, in the absence of a proper test dataset for the validation of classification results, rules' uncertainty values can be applied to represent their corresponding hit ratio. The collection of reliable ground truths for validation purposes is usually an expensive task in terms of time and economic costs (Bruzzone and Prieto, 2001). Consequently, in many cases, it may not be possible to rely on test data to ensure good performance of a classifier. Accordingly, aside from using traditional accuracy metrics as a single number derived from a confusion matrix, mapping and thresholding the rule-exclusive hit ratio in a classification scheme is worthwhile for visualising general patterns of high and low accuracy values within the classified map and quantifying the accuracy of prediction in specific targeted locations.

3.4.4. Limitations of DoTRules and future work

Although the results obtained by DoTRules are encouraging, further comparative experiments with additional hyperspectral imagery datasets should be implemented. This can be more useful with a particular focus on assessing the classification performance at higher levels of disaggregation, such as a class-level accuracy assessment. As some of the required parameters for the DoTRules implementation are subjective, such as 1) the rule uncertainty threshold, 2) the minimum and maximum length of random rules and 3) the optimum number of rule sets, more research may be beneficial in the computational optimisation of DoTRules parameters. Our further work is focusing on the development of more computationally efficient schemes for the ensemble framework.

Another limitation of the proposed ensemble framework is the fact that the proposed framework is less efficient for very low sample sizes (i.e., 1% or less). DoTrules usually needs a larger training set to extract the underlying relationships between variables. This is a common requirement for all ensemble methods except RoF. Although RoF is the best performing algorithm for the 1% sample size of Indian Pines, it benefits from the transformation of the hyperspectral data.

3.5. Conclusion

We have applied DoTRules—a Dictionary of Trusted Rules—as an innovative ensemble framework for classifying hyperspectral data with high accuracy estimates compared with other popular classification algorithms. DoTRules' classification accuracy was superior to six other popular and state-of-the-art ensemble and non-ensemble algorithms. In the case of DoTRules, every rule within any rule set can be accessed, and their corresponding uncertainty value may be observed. This feature is unique to DoTRules and the absence of this ability underpins a common criticism of many ensemble algorithms (including many of the algorithms applied here) as black-box classifiers. Furthermore, DoTRules is also capable of quantifying and mapping the uncertainty of these classification rules, prior to the image classification where the uncertainty values can be applied as an estimate of the hit ratio. While the entropy product of DoTRules provides spatial insights, including the location of less reliable classification rules as well as more reliable ones, regardless of the test sample dataset availability, it can also certify and locate less accurate rules using the estimated hit ratio. The spatial exploration of rule uncertainty in hyperspectral image classification is beneficial for the early prediction of success or failure of a classifier in specific geographic locations. The uncertainty maps may also serve to enhance the application of map products by alerting map users to the spatial variation of rules' hit ratio over the entire mapped region. This, together with the simplicity and accuracy of DoTRules, indicates that the methodology offers new features and is ready for operational use by the remote sensing community.

4. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change

The aim of chapter 4 is to apply DoTRules—a dictionary of trusted rules—as a transparent alternative to calculate transition potential in cellular automata models. The method applies Shannon entropy to assess the uncertainty of each rule, and the most trusted rules are used to project future LUCC. This chapter corresponds to objective three - modelling uncertainty Type A and has been published in **Environmental Modelling & Software** journal on 27th December 2018.

Roodposhti, M.S., Aryal, J. and Bryan, B.A., 2019. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. Environmental Modelling & Software, 112: 70-81.

Abstract: Despite recent advances in quantifying land-use/cover change (LUCC) transition potentials, transition rules are often not transparent and uncertainty is rarely made explicit. Here, we introduce DoTRules—a dictionary of trusted rules—as a transparent alternative to calculate transition potential in cellular automata models. Rules relate LUCC variables to the observed historical changes. Shannon entropy is calculated to assess the uncertainty of each rule, and the most trusted rules are used to project future LUCC. DoTRules produces rule-level uncertainty estimates, which can be mapped. In a case study of the Ahvaz region of Iran, the overall accuracy of LUCC simulation calibrated using DoTRules approach was very similar to simulations calibrated with the state-of-the-art random forest, but DoTRules provides a transparent approach where transition rule information and uncertainty can be readily accessed and interpreted. The results demonstrate that DoTRules has potential to derive new insights into LUCC processes.

4.1. Introduction

Cellular automata (CA) were conceptually established by John von Neumann (1903–1957) during the 1950s. Due to their simplicity and capacity to simulate spatial patterns, CA have rapidly gained popularity as a tool for modelling spatial dynamics of many environmental phenomena such as plant population dynamics (Xu et al., 2010), forest fire spread (Zheng et al., 2017), slope failure (Liucci et al., 2017), debris flow (D'Ambrosio et al., 2003; D'Ambrosio et al., 2006), urban sprawl (Mustafa et al., 2018a; Van Vliet et al., 2009), land-use/cover change (LUCC) (Hewitt and Díaz-Pacheco, 2017; Hewitt et al., 2014; White and Engelen, 1997) and more. Though cellular automata can handle very complex spatial situations for modelling environmental phenomena, their conceptual basis is straightforward. A cellular automaton consists of a large number of cells, which can change their state according to specific rules. In many applications such as modelling fire spread (i.e. bushfire or forest fire), urban sprawl modelling and specially LUCC simulation, a set of neighbourhood and suitability values are defined reflecting the influence of external factors affecting the state transitions for each cell. Finally, there is a set of rules defining transition potential of a cell from one state to another.

In terms of LUCC models, the transition demand and the transition potential are typically the two main requirements (White and Engelen, 1993) for model implementation using cellular automata. First, historical rates of land-use change are calculated which are used to calibrate the total amount of land-use change occurring in each time step. This is termed transition demand. Second, the spatially explicit probabilities of land-use change, or transition potential, are calculated. Transition potential represents the behavioural propensities of the actors determining land-use change and is defined based on the inferred logic from a set of transition rules. Transition rules are general structures that offer an easily understandable and transparent way to find the most reliable class allocation (Russell et al., 2003). In practice, transition rules capture the relationships between land-use and a suite of independent predictor variables. The effectiveness of cellular automata land-use models in informing land-use planning depends upon the efficient extraction of reliable and transparent transition rules (Han et al., 2015; Hewitt et al., 2014).

Numerous machine-learning and statistical methods have been used to calculate land-use transition rules and map transition potential for use in cellular automata land-use models (Basse et al., 2014; Berberoğlu et al., 2016; Clarke et al., 1997; Ku, 2016; Liu et al., 2014; Mustafa et al., 2017; Mustafa et al., 2018a; Mustafa et al., 2018b). Methods frequently applied include association rule learning (Al-kheder et al., 2008; Liu and Jiang, 2011), artificial neural network (Basse et al., 2014; Li et al., 2013b), maximum margin (Rienow and Goetzke, 2015; Yang et al., 2008), instance-based (Castilla and Blas, 2008; Li et al., 2015), regression (Ku, 2016; Long et al., 2014), decision tree (Ballestores Jr and Qiu, 2012; Basse et al., 2016), and probabilistic (Arsanjani et al., 2011; Vaz et al., 2015) methods. Others such as evolutionary, deep learning, reinforcement learning, dimensionality reduction, Bayesian, and regularisation methods have been used less frequently (Kamusoko and Gamba, 2015; Li et al., 2015; Verstegen et al., 2014; Zhang et al., 2008). Each of these methods employs structurally different numerical formulations, which affect the accuracy and transparency of automata-based LUCC models. Few of these methods facilitate the transparent extraction of transition rules and their corresponding uncertainty.

Transparent transition rules enable both an enhanced ability to appreciate the relationships between land-use change (or other similar environmental phenomena) and predictor variables. This is necessary to understand model structure. Hence, the inferred logic of model structure captured in transition rules can be visualised, dissected, and deciphered, and can be generalized and applied to address similar problems. This can be invaluable for informing error checking and enabling model validation. The explicit identification of transition rules can also help understand the nature of the major land-use transitions (Tayyebi et al., 2014). Incomplete information about transition rules and their uncertainty may impede the understanding of land-use change processes (Koomen and Borsboom-van Beurden, 2012; Pozoukidou, 2005). However, many approaches have been subject to limitations in their ability to clearly identify transition rules and have been criticised as being black-boxes (Islam et al., 2018; Li and Yeh, 2002; Qiu and Jensen, 2004; Waddell, 2002). For instance, Kamusoko and Gamba (2015) performed and compared cellular automata calibrated using random forest, support vector machine, and logistic regression where the performance of the random forest model was attributed to the relatively accurate transition potential maps. However, apart from logistic regression which is capable of revealing the relative global contribution of each variable to the land-use change process, none of these methods is capable of implementing accessible and transparent sets of transition rules at the pixel level. Transparent approaches are more desirable than black-box approaches for transition rule detection, even if this preference trades-off some performance (Tseng et al., 2008; Uzuner et al., 2009). This includes, but is not limited to, LUCC plans where it is required to provide insight into the internal decision-making process of algorithms for a better interpretation of the result. Similar requirements also exist for other environmental applications aimed at improving the quality of management plans in the context of natural hazards (Lai et al., 2016; Royston et al., 2012; Shadman Roodposhti et al., 2016), water treatment (Gibert et al., 2010), soil erosion (Adinarayana et al., 1999), and farming systems (Moore et al., 2014).

Simulated change and persistence in land-use patterns need to be interpreted and validated via a better understanding of uncertainty both at the rule level and spatially. While a few studies have successfully mapped the spatial distribution of classification uncertainty (Bryan et al., 2009; Khatami et al., 2017), providing estimates of classification uncertainty at rule-level (i.e. for each rule) may also provide complementary insights into land-use change processes. However, rather than providing uncertainty estimates, land-use modelling studies typically report on the accuracy of LUCC analyses using global methods such as confusion matrices and the Kappa index (Congalton, 1991). Confusion matrices are usually calculated to allow for global measures of accuracy (i.e. overall accuracy) to be generated and lack the ability to quantify the spatial distribution of classification accuracy (Tsutsumida and Comber, 2015). Similarly, the Kappa index does not consider the disagreement in the classification accuracy (Pontius Jr and Millones, 2011; Stein et al., 2005a). Global measures of accuracy require ground truth data while in terms of land-use change models it is a time-consuming process to prepare land-use ground truth maps for every simulated map of land-use. In addition, there is no ground truth for future

land-use scenarios and therefore these methods provide no basis for quantifying confidence in simulated future land-use maps. Explicit rule-level uncertainty estimates can assist land-use planners and policymakers to understand the uncertainties associated with major land-use transitions. The mapping of these uncertainties can identify locations where simulated land-use allocation occurs with high confidence or areas of low confidence, which can both be useful in land-use planning.

Here, we develop a new algorithm, DoTRules—Dictionary of Trusted Rules—for modelling land-use transition potential for application in automata-based LUCC models, which features the transparent identification of transition rules and quantifies their uncertainty. It also enables the mapping of corresponding land-use transition uncertainties. In DoTRules, the uncertainty of transition rules is quantified using Shannon entropy. Dissecting transition rules and their corresponding uncertainty enables the better understanding of the core rules governing major land-use/cover dynamics, which is useful for informing land-use planning. We also show that the uncertainty values can be applied as an approximation of simulation accuracy. We describe the DoTRules algorithm and demonstrate its application to the Ahvaz region, Iran. We quantify the uncertainty of LUCC simulation calibrated using DoTRules to calculate land-use transition rules and compare the results with simulations based on random forest transition rule detection. We discuss the advantages and disadvantages of the new approach for LUCC modelling more generally and the application of DoTRules in the calculation of transition potential maps for LUCC models to broader environmental processes where it is necessary to understand the direction and magnitude of state transitions.

4.2. Description of DoTRules

DoTRules is a moderate-speed rule-based algorithm for calculating transition potential in LUCC according to a dictionary of trusted rules where categorical/discrete data are involved. It is similar to the random forest algorithm (Breiman, 1996; Breiman, 2001) insofar as rule sets are used to select the mode response (i.e. most frequent land-use class) among every available potential response variable. However, instead of generating random trees, DoTRules operates by constructing many transition rules within various rule sets derived from a training dataset, with land-use assigned to the most frequently occurring class. The rule construction process is fulfilled using concatenation of discrete predictor variables and the entropy of each rule is then calculated as an estimate of accuracy. The DoTRules procedure was implemented in R (R Core Team, 2017) and consists of the following six steps.

STEP 1: Assembling the data.

Training data is represented by a set of grid cells $I=\{ i_1, i_2, \ldots, i_m \}$. Each grid cell i in I has a value x_{ij} for each of the independent predictor variables or *criteria* $J=\{j_1, j_2, \ldots, j_n\}$ where in this study there are nine independent predictor variables (Table 4.1). Criteria are discretised predictor variables which can be derived from either native categorical data (e.g. land-use class) or classified continuous data (e.g. distance to road). Thus, for each criterion, x_{ij} can adopt one of a fixed set of possible classes specific to that criterion, which we represent as the set H for every j in J (Table 4.1). Note that each criterion j will have a different set of classes H but, for clarity, here we do not index H by j. Each grid cell i has a corresponding land-use class l_i which are also discrete semantic attributes from the set of five land-uses L\{u: urban, a: agriculture, b: bare lands, r: roads and w: water bodies\}.

STEP 2: Calculating Shannon entropy and prioritising criteria.

For each criterion j in J, we calculated the frequency of grid cells i within each criterion class h in H occurring within each land-use class l in L, represented as $P_{l,h,j}$:

$$P_{l,h,j} = \frac{\sum_{i \in I} [x_{ij} = h][l_i = l]}{\sum_{i \in I} [x_{ij} = h]} \quad \text{for } \forall\ l \text{ in } L,\ h \text{ in } H,\ \text{and } j \text{ in } J \tag{4.1}$$

Note that […] are Iverson brackets where [Q] (quantity) equals 1 if true, and 0 if false. The term $\sum_{i \in I} [x_{ij} = h]$ is the number of grid cells in criterion class h.

In information theory, entropy is the quantitative measure of system disorder, instability, and uncertainty (Shannon, 2001). The Shannon entropy is the quantitative measure of uncertainty in this

study. Here, we calculate the entropy of land-use class occurrences within each criteria class h across all criteria j.

$$e_{hj} = -\sum_{l \in L} P_{l,h,j} \ln P_{l,h,j} \tag{4.2}$$

The entropy of each criterion was then calculated as the average entropy of its classes $e_{h,j}$ weighted by the proportion of cells in each class:

$$e_j = \sum_{h \in H} e_{h,j} \sum_{i \in I} [x_{ij} = h] / |I| \tag{4.3}$$

where $|I|$ is the set of grid cells in the training dataset. The criteria were then ranked and prioritised according to their average entropy e_j with higher priority criteria being those with the lower entropy, represented by the ordered set of criteria priority J'.

STEP 3: Creating a rule set.

We then concatenate grid cell criteria values x_{ij} as per criteria priority J' in order to form a rule set D. The concatenation of two or more characters is the string formed by them in a series (i.e. the concatenation of 12, A7, and 5$ is 12A75$). Equation 4 illustrates the grid cell values for criteria ranked in order of priority (i.e. lowest entropy) concatenated for each grid cell (row) i, thereby creating a unique rule for each grid cell in the training dataset.

$$D = \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{i1} \\ x_{|I|1} \end{pmatrix} \| \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{i2} \\ x_{|I|2} \end{pmatrix} \| \cdots \| \begin{pmatrix} x_{1j'} \\ x_{2j'} \\ \vdots \\ x_{ij'} \\ x_{|I|j'} \end{pmatrix} \| \begin{pmatrix} x_{1|J'|} \\ x_{2|J'|} \\ \vdots \\ x_{i|J'|} \\ x_{|I||J'|} \end{pmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j'} & x_{1|J'|} \\ x_{21} & x_{22} & \cdots & x_{2j'} & x_{2|J'|} \\ \vdots & \vdots & & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij'} & x_{i|J'|} \\ x_{|I|1} & x_{|I|2} & \cdots & x_{|I|j'} & x_{|I||J'|} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_i \\ d_{|I|} \end{bmatrix} \tag{4.4}$$

Note that following the concatenation and extraction of rules, every rule within the dictionary has maintained its single land-use class $l_i \in L$. We then aggregate duplicate rules where grid cells have exactly the same values for all criteria, leaving a parsimonious new rule set of unique rules D' derived by aggregating D. The frequency of occurrence of all potential land-use classes l in L is then calculated for each unique rule d' in D':

$$\begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_{|D'|} \end{bmatrix} \begin{matrix} \rightarrow \\ \rightarrow \\ \vdots \\ \rightarrow \end{matrix} \begin{bmatrix} f(l_{1u}) & f(l_{1a}) & f(l_{1b}) & f(l_{1r}) & f(l_{1w}) \\ f(l_{2u}) & f(l_{2a}) & f(l_{2b}) & f(l_{2r}) & f(l_{2w}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f(l_{|D'|u}) & f(l_{|D'|a}) & f(l_{|D'|b}) & f(l_{|D'|r}) & f(l_{|D'|w}) \end{bmatrix} \tag{4.5}$$

The land-use class (i.e. u, a, b, r, w from set L) with the highest frequency (i.e. the mode) is then assigned to each corresponding unique rule d'.

STEP 4: Calculating and mapping the uncertainty of land-use prediction.

Considering every unique rule d' from our rule set D', a Shannon entropy value is then calculated based on the frequencies of each possible land-use class (Equation 4.5) using Equation 4.2. This can inform both the spatial distribution of uncertainty in land-use predictions and provides transparent transition rules for informing land-use planning. The spatial distribution of uncertainty is quantified and mapped by the entropy of each unique rule back to the grid cells corresponding to each rule. Each grid cell is then allocated to the land-use class with the highest frequency for its corresponding rule.

STEP 5: Classify land-use of test dataset according to the dictionary of trusted rules.

Above we describe the process of creating the dictionary of trusted rules and allocating the most likely land-use class for each rule based on frequency. The land-use class can now be predicted for the rest of the study area dataset (i.e. the test dataset). To do this, we follow the same procedure to set up a rule set for the test dataset. We then match each test data rule with its equivalent in the dictionary of trusted rules using many to one matching (i.e. matching many rules concatenated from test grid cells to individual trusted rules calculated using the training dataset) and allocate the most likely land-use to each test data rule. This can then be mapped back to the grid cell level as each rule in the test dataset corresponds to a grid cell.

STEP 6: Handling null values

Finally, as there is always a possibility of encountering 'null' values using the DoTRules approach where new cells in the test dataset present combinations of criteria states not encountered in the training data. Here, using the same training and test sample, we will sequentially exclude the least informative (i.e. highest entropy and lowest ranked) independent predictor variables or *criteria J'* from our analysis and re-execute DoTRules step 3 to 5. This generates a second rule set (i.e. a sub-rule set) which contains fewer unique rules *d'* (in the corresponding *D')* and 'null' records as it contains fewer criteria classes. We then repeat in developing some new sub-rule sets until all 'null' values in our primary rule set are covered by some corresponding sub-rules among secondary sub-rule sets. The best sub-rule to replace a null-matching rule in our test rule set is the one with the lowest entropy while those rules with higher entropy values (higher than the specified threshold) are eliminated.

4.3. Methods

4.3.1. Study area

The study area was the Ahvaz region of south-west Iran. Ahvaz city is the capital and largest city of Khuzestan province (Figure 4.1). The population of Ahvaz increased from 334,399 to 1,338,126, from 1976 to 2015, with attendant growth in urban areas. The Karun River, 850 km long and Iran's largest, splits the city into western and eastern parts then joins the Arvand Rood River and continues toward the Persian Gulf. The climate is semi-arid, with a mean annual precipitation of 252 mm and an average annual temperature of 26.9°C. June is the driest and warmest month and January is wettest and coolest.

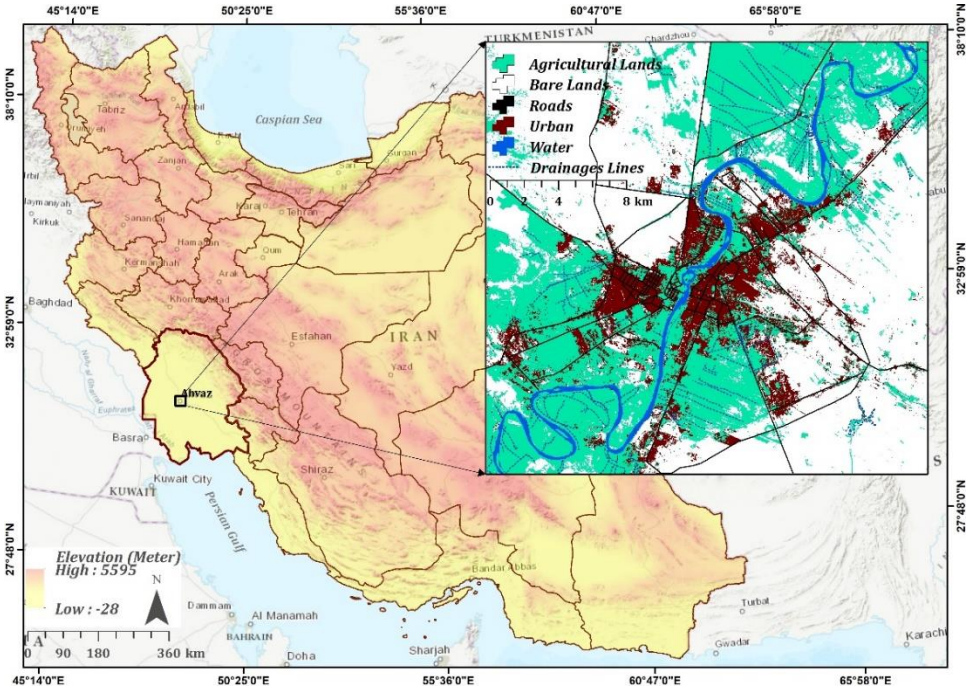


Figure 4.1. Location and land-use in the study area of Ahvaz, Iran.

The major land-use transition trends in the study area included the rapid growth of built-up urban areas (transformed from bare lands and agricultural lands) and agricultural lands (transformed from bare lands) from 1985 up to 2015. A minor transition from agricultural lands to bare lands was also observed. Urban growth was driven by rapid population growth. Agricultural production has increased to meet increasing local demand, supported by the abundance of water and fertile soils. Various agricultural commodities are produced such as wheat, barley, oilseeds, rice, sugar cane, medicinal herbs, as well as orchard crops such as palm, citrus, and olives (Rangzan et al., 2008).

4.3.2. LUCC simulation process overview

LUCC simulation was implemented in five Phases (Figure 4.2). Phase 1 involved data collection, preparation, and pre-processing, including land-use/cover classification, undertaken using geospatial analysis software (ENVI and ArcGIS). Neighbourhood analysis, cost-distance layer preparation, and data discretisation was then done using the raster package (Hijmans and van Etten, 2014) in R. Phase 2 involved calculating rates of land-use change between 1985-2000 and identifying major land-use change transitions for specifying land-use change demand (Islam et al., 2018; Kamusoko and Gamba, 2015).

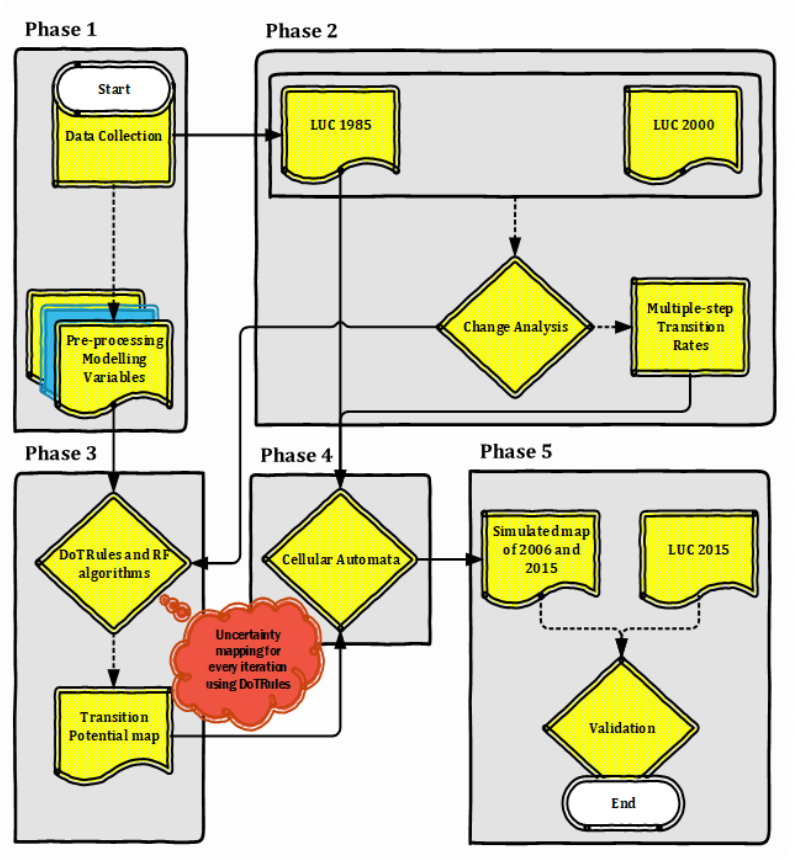


Figure 4.2. Schematic representation of methodology implementation in five phases.

Phase 3 involved the use of DoTRules and random forest (RF) algorithms to calculate LUCC transition potential maps using a training sample of randomly selected grid cells. In Phase 4, the land-use map of 1985, transition potential maps, and the estimated rates of land-use change for the primary land-use classes were integrated into a cellular automata model in R. The CA model was calibrated to apply 30 annual iterations (one for each year 1985 – 2015). Finally, in Phase 5, the predictive accuracy of the simulated land-use maps for 2015 was validated against the classified land-use map for the same year using 100,000 random points for three simulated land-use classes (i.e. urban, agriculture and bare lands). Finally, we then compared the accuracy of cellular automata land-use change models calibrated using DoTRules and RF algorithms.

4.3.3. LUCC modelling variables and data sources

To analyse the trend of change and calculate transition potential maps, which are required for simulating future LUCC, Landsat images of the years 1985, 2000 and 2015 were used. The two earlier Landsat images (i.e. 1985, 2000) were used for land-use change analysis and the calculation of land-use transition potential maps while the Landsat image for 2015 was used for validation. Five groups of variables including cell state (CS), neighbourhood variables (NV), suitability variables (SV), a target

variable, and validation data were extracted from the main data sources for LUCC simulation (Table 4.).

Table 4.1. Applied variables including cell state (CS), neighbourhood value (NV), suitability value (SV), target variable and validation data along with a description, units, data source and number of classes (H).

Variable type	Description	Unit	Source	Classes (H)
Cell state (CS)	Land-use state of cell *x,y* at times 1985	-	Landsat	5
Neighbourhood variables (NV)				
u_{xy}^{t}	Number of neighbouring urban cells for cell *x,y* (u_{xy}^{t}≤8)	-	Landsat	9
a_{xy}^{t}	Number of neighbouring agricultural cells for cell *x,y* (a_{ij}^{t}≤8)	-	Landsat	9
b_{xy}^{t}	Number of neighbouring bare lands cells for cell *x,y* (b_{xy}^{t}≤8)	-	Landsat	9
r_{xy}^{t}	Number of neighbouring road cells for cell *x,y* (r_{xy}^{t}≤8)	-	Landsat	9
w_{xy}^{t}	Number of neighbouring water cells for cell *x,y* (w_{xy}^{t}≤8)	-	Landsat	9
Suitability variables (SV)				
D_{xy}^{t}	Distance from drainage networks for cell *x,y* at time *t*	metre	Google Earth	4
R_{xy}^{t}	Distance from road networks for cell *x,y* at time *t*	metre	Google Earth	4
s_{xy}^{t}	Slope value of cell *x,y* at time *t*	degrees	SRTM DEM	10
U_{xy}^{t}	Distance from urban edge for cell *x,y* at time *t*	metre	Landsat	10
Target variable	Land-use state of cell *x,y* at time 2000	-	Landsat	5
Validation data	Land-use state of cell *x,y* at time 2015	-	Landsat	5

Note that for the NV, as neighbourhood configuration is known to affect cellular automata simulation (Fuglsang et al., 2013; Lauf et al., 2012; Verstegen et al., 2014; White and Engelen, 1993), the optimal kernel size k=8 for neighbourhood analysis was selected based on initial cross-validation. All variables were resampled to the 30-metre grid cell resolution of the Landsat data, totalling 734,328 cells across the study area. The derivation and use of these variables is described in detail below.

4.3.3.1. Landsat archive and image classification

Image pre-processing involved normalization for the region of interest. Land-use/cover maps of the study area were then classified using a support vector machine classifier with geospatial analysis software (ENVI and ArcGIS), achieving an overall accuracy of ≥85%. During this process, all grid cells were allocated to one of five land-use/cover classes: urban area, agricultural land, bare land, roads, and water bodies for the 1985, 2000 and 2015 images (Figure 4.3).

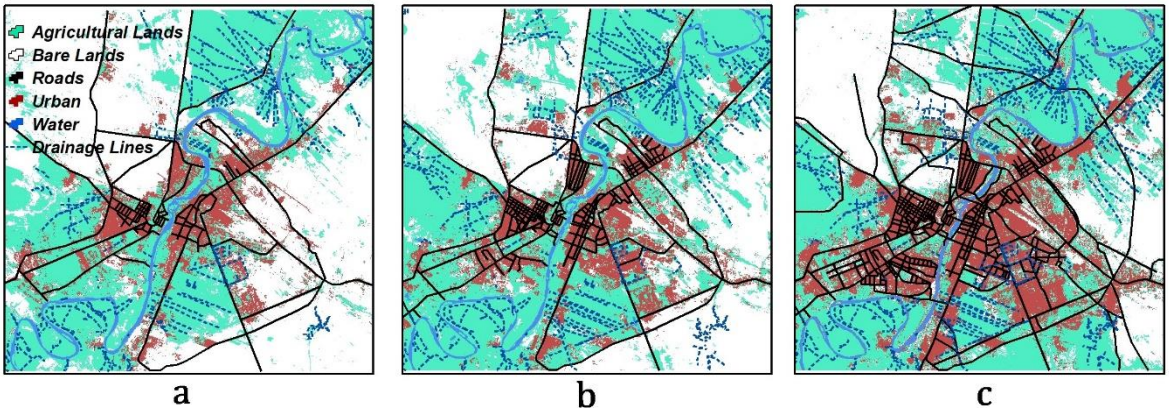


Figure 4.3. Landsat-derived land-use/cover maps of the study area for (a) 1985, (b) 2000 and (c) 2015.

4.3.4 Land-use change analysis to compute transition demand

The rate of simulated land-use change, or transition demand, in cellular automata models needs to be calibrated to observed rates by quantifying the historical amount of change for each land-use type (Hewitt and Díaz-Pacheco, 2017; Kamusoko et al., 2009; Kamusoko and Gamba, 2015; Pastor et al., 1991). We calculated transition demand based on the Landsat-derived land use/cover maps for 1985 and 2000 (Figure 4.3). The type and frequency of land-use change between 1985 and 2000 were cross-tabulated. The time interval used for calibration for the 1985–2000 transition matrix was 15 years. Land-use transition probabilities were calculated as average annual rates of change following previous studies (Hewitt and Díaz-Pacheco, 2017) in order to take account of annual change demand for each cellular automata iteration.

4.3.5 Computation of land use/cover transition potential maps

In constructing a training dataset for calculating transition potential maps, we selected 300,000 grid cells randomly from the major 1985–2000 land-use change categories of "bare lands to urban", "agriculture to urban", "bare lands to agriculture", "agriculture to bare lands" and "no change" (that is, areas which remained unchanged as urban, agriculture, bare lands and water bodies). Independent predictor variables were then calculated for the initial year t = 1985 including the cell state (i.e. land-use map), along with neighbourhood and suitability variables (Table 4.1). We then computed the transition potential map for 1985 using DoTRules and the random forest algorithm as implemented in the randomForest package (Liaw and Wiener, 2002) available in R. CS, NV, and SV were recalculated each year based on the simulated land-use and used to update the land-use transition potential map each year.

4.3.6 CA-based land-use change simulation

In traditional cellular automata models, the evolution of the future cell state is determined by the following formula (Al-shalabi et al., 2013; Martinez et al., 2012; Wu, 1998):

$$S_{xy}^{t+1} = f(S_{xy}^{t}, \rho_{xy}^{t}, \omega_{xy}^{t}) \quad (6)$$

where, S_{xy}^{t} represents the land-use state for a cell at location (x,y) at time t. ρ_{xy}^{t} is composed of a set of suitability measures for the cell at time t. ω_{xy}^{t} is the state of neighbouring cells at time t, and f is a transition function.

Three datasets, (1) the initial land use/cover map (1985); (2) the transition potential maps (1985–2000); and (3) the transition demand, were used to simulate land use/cover up to 2015 using a cellular automata (Hewitt et al., 2013). Transition demand, calculated via the land-use change analysis, determined the amount of land-use change in each simulated year, while the land-use transition potential determined the location and type of change (Yang et al., 2016). For each simulation year, land-use was allocated by finding the grid cell with the maximum transition potential. If the new land-use was less than that demanded then the change was made. The cell with the next highest transition potential was then found and the change made if the new land-use was less than that demanded. This process was repeated until all land-use demands were met for that year (Yang et al., 2016). This whole process was repeated for 30 annual iterations to simulate land-use from 1985 to 2015 where a new land-use map is produced at the end of each iteration.

4.3.7. Comparing DoTRules with random forests

To quantify DoTRules' performance in calculating land-use transition potential, we implemented the same CA-based LUCC simulation scheme but using the RF algorithm (Kamusoko and Gamba, 2015) to calculate transition rules. The RF algorithm provides an appropriate benchmark for assessing the performance of the DoTRules scheme because of the high performance typically found in predictive modelling (Caruana and Niculescu-Mizil, 2006). RF is also computationally efficient and suitable for

large training data (Mahapatra, 2014). We compared the overall accuracy of both DoTRules and RF-based cellular automata simulation of LUCC for the Ahvaz study area.

4.4. Results

4.4.1. Variable importance and transition rules

A major product of DoTRules is calculated Shannon entropy values, which are used for prioritising criteria before assembling transition rules. Criteria are ranked and prioritised according to their average entropy e_j (eq. 4.3) with higher priority criteria being those with the lower entropy (Table 4.2). In our case study, 24,437 transition rules were assembled and applied for the purpose of LUCC simulation. Here, every single rule/sub-rule is defined by a unique string which represents criteria values, along with a frequency distribution of potential land-use class labels, rule-exclusive hit ratio, rule-exclusive entropy value, and mode land-use label (Figure 4.4).

Table 4.2. Sorted variables of LUCC model from Table 1 along with their priority value (mean entropy e_j).

Variable	CS	b^t_{xy}	U^t_{xy}	a^t_{xy}	R^t_{xy}	u^t_{xy}	D^t_{xy}	w^t_{xy}	r^t_{xy}	s^t_{xy}
Priority	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
e_j	0.815	0.956	1.001	1.005	1.009	1.028	1.070	1.099	1.113	1.142

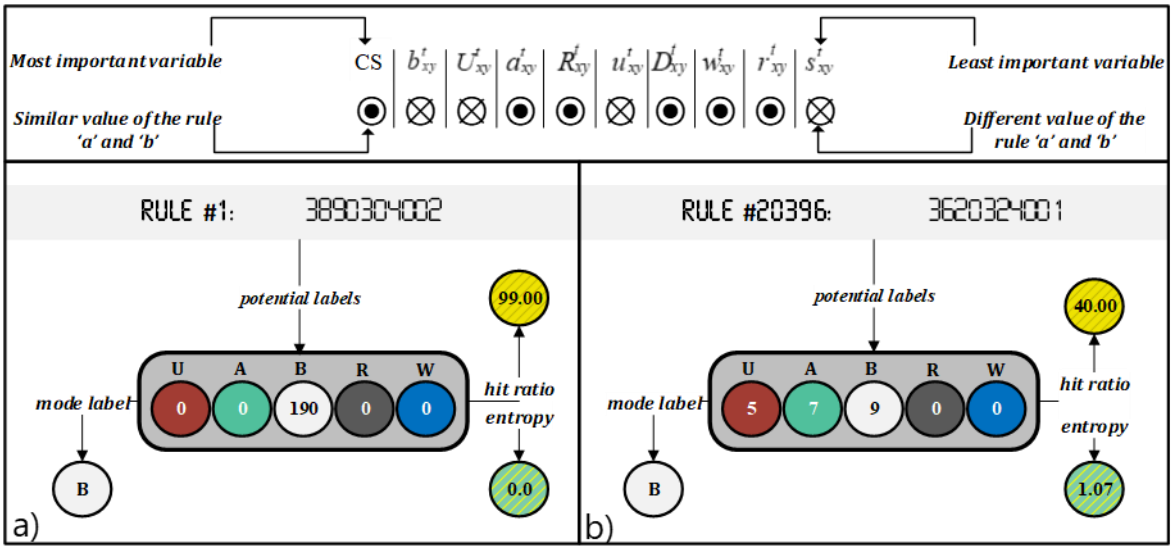


Figure 4.4. Samples of low uncertainty (a, left) and high uncertainty (b, right) rules extracted using DoTRules. The string of numbers highlighted in grey is represents concatenated class labels from the 10 variables from Table 4.1 prioritised by their predictive ability as shown in Table 4.2.

The transparency of DoTRules opens up information contained in the transition rules for critical observation or examination. For example, retained information in the following rules of Figure 4.4 can be dissected and retrieved once the variable priority is clarified. Here, both rules have the same value for cell state, neighbouring agricultural cells, distance to road, distance to drainage, and neighbouring water and road cells. However, the rules have different values for neighbouring bare land cells, distance from urban, neighbouring urban cells, and slope. The two rules have the same mode label but different uncertainties and hit ratios.

4.4.2. Simulation performance

Following the simulation procedure based on the transition potential maps as calculated by DoTRules and RF, the results are mapped and then validated by a 100,000 validation test points of 2015 land-use data map (Figure 4.5).

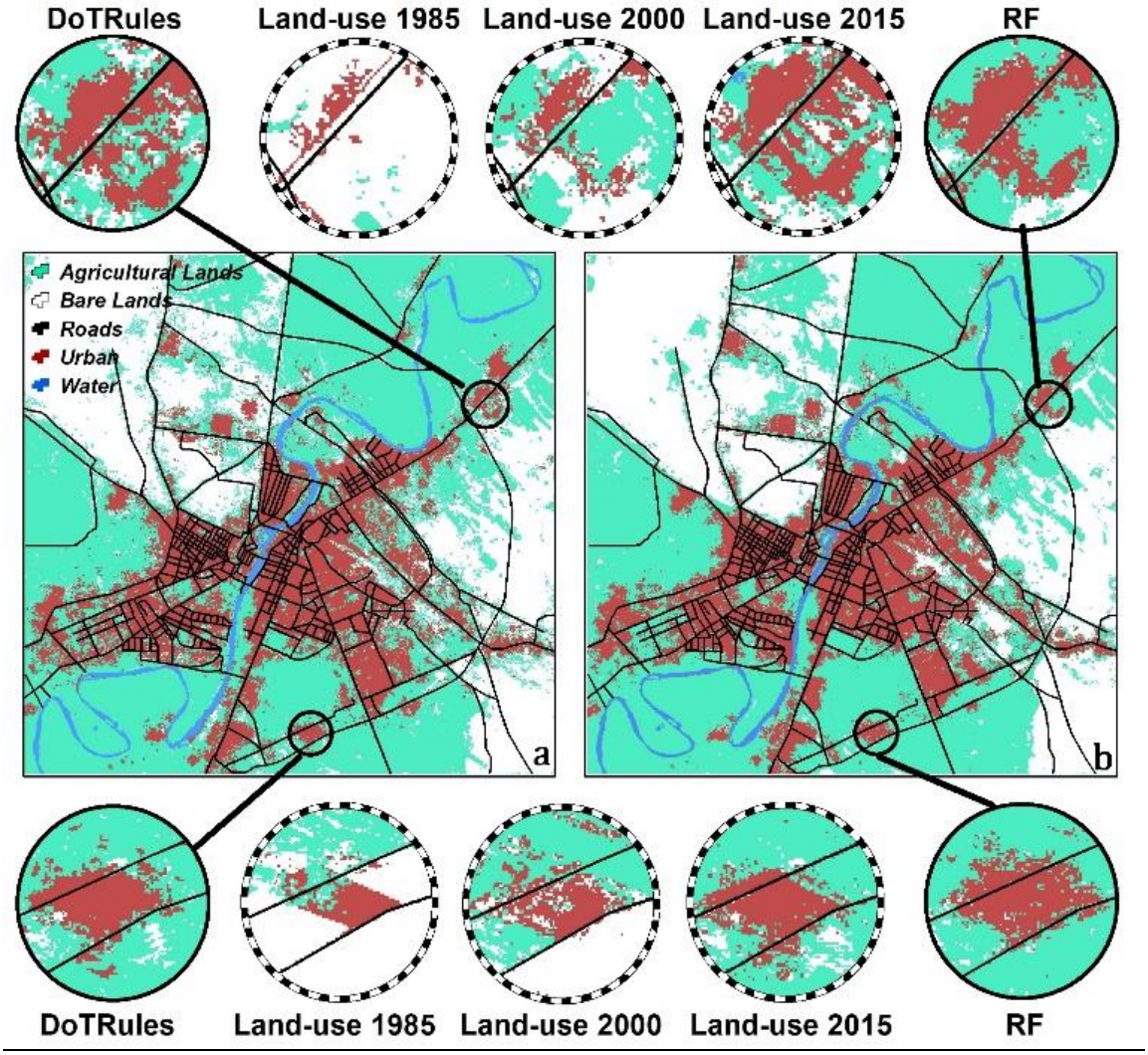


Figure 4.5. Simulated land-use/cover map of Ahvaz for the year 2015 using DoTRules (a) and random forest (b). The sub-plots demonstrate the local differences of the two algorithms against validation data. Dashed border represents Landsat-derived land-use maps for that year.

This comparison is done only for simulated land-use classes including urban, agriculture and bare lands. Considering the fact that land-use map of 2015 was not involved in the preparation of transition potential maps, the overall accuracy of the LUCC simulation using DoTRules (75.4) was very similar to that based on RF (75.8). Although both algorithms demonstrate broad spatial similarities, LUCC simulation results of the Ahvaz study area for the target year of 2015 also display localised differences. Land-use simulation of both DoTRules and RF were promising in identifying, retaining, and preserving the spatial details of the 2015 land-use/cover maps.

4.4.3. Uncertainty of major land-use transitions

A key advantage of DoTRules is that all major transitions can be identified and dissected, enabling the analysis of major trends of change/persistence along with their corresponding information such as uncertainty and frequency (Table 4.3). In terms of land-use change/persistence transitions and their uncertainty values, 'bare lands to bare lands' persistence had the lowest transition rule uncertainty, while 'agriculture to urban' change was the most uncertain land-use transition. Transition rules indicating land-use persistence tended to have lower uncertainty than did rules indicating a change from one land-use to another.

Table 4.3. Estimated frequency and entropy of major land-use transitions.

From-To	Frequency	Average entropy	Uncertainty of rules in percent of total		
			e < 0.2	**0.2 ≤ e ≤ 0.6**	**e > 0.6**
Urban-Urban	2218	0.483	43.61	11.21	44.14
Agriculture-Agriculture	4293	0.491	40.15	12.50	47.35
Agriculture-Urban	3028	0.351	31.03	9.67	59.30
Agriculture-Bare lands	3780	0.542	36.84	11.68	51.41
Bare lands-Bare lands	3539	0.341	53.11	11.6	34.67
Bare lands-Agriculture	2108	0.361	39.36	12.47	48.10
Bare lands-Urban	1794	0.387	30.59	13.71	55.67

4.4.4. Rule-level spatial uncertainty

The DoTRules spatial uncertainty product can facilitate the better understanding of uncertainty of transition rules in the mapped land-use predictions. As there will be one uncertainty map for every simulation year, it helps to understand where LUCC simulation results are less or more reliable for each iteration. Considering the results of LUCC simulation using DoTRules, the mean uncertainty maps (for 30 iterations) demonstrates a large extent of low (L), very low (VL) and extremely low (EL) uncertainty classes (i.e. class labels) characterised by a low uncertainty estimate may be observed within the inner boundaries, where LUCC is less active. Patches of high (H), very high (VH) and extremely high (EH) uncertainty occurred where LUCC is more active, particularly those areas located at the interface between urban and agricultural lands (Figure 4.6).

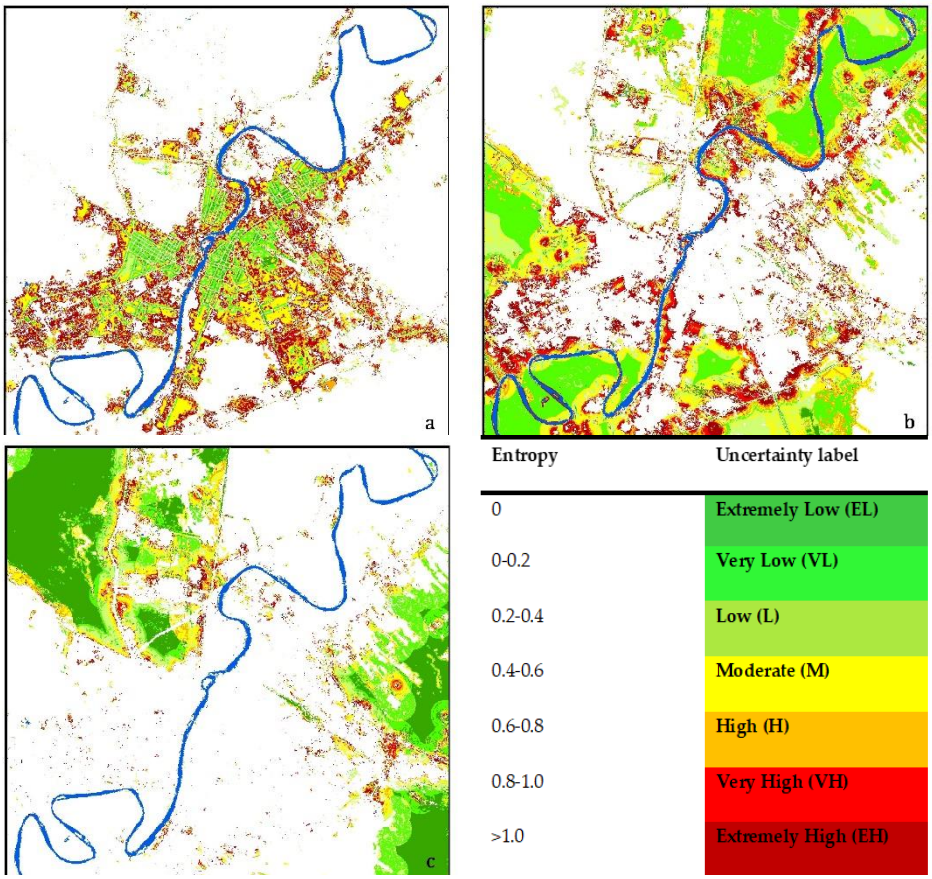


Figure 4.6. Uncertainty map of DoTRules for three major land-use classes including (a) urban, (b) agriculture, (c) bare lands, during 30 years of simulation (up to 2015) in the study area.

4.4.5. Uncertainty and accuracy

In this study, 24,437 unique rules were detected for the primary rule set where their relevant hit ratio is measured using available test data of 2015. Uncertainty was closely related to hit ratio (i.e. percent of correctly assigned land-use labels for a rule) of transition rules where the hit ratio of transitions exponentially decreases with increasing uncertainty (R2= 0.89). Thus, low uncertainty transitions are associated with a higher hit ratio (Figure 4.7). Lower uncertainty means there is an obvious land-use class (mode land-use class label) for a rule, while high uncertainty reflects that there are several candidate land-use classes for that rule, which results in less accurate land-use prediction.

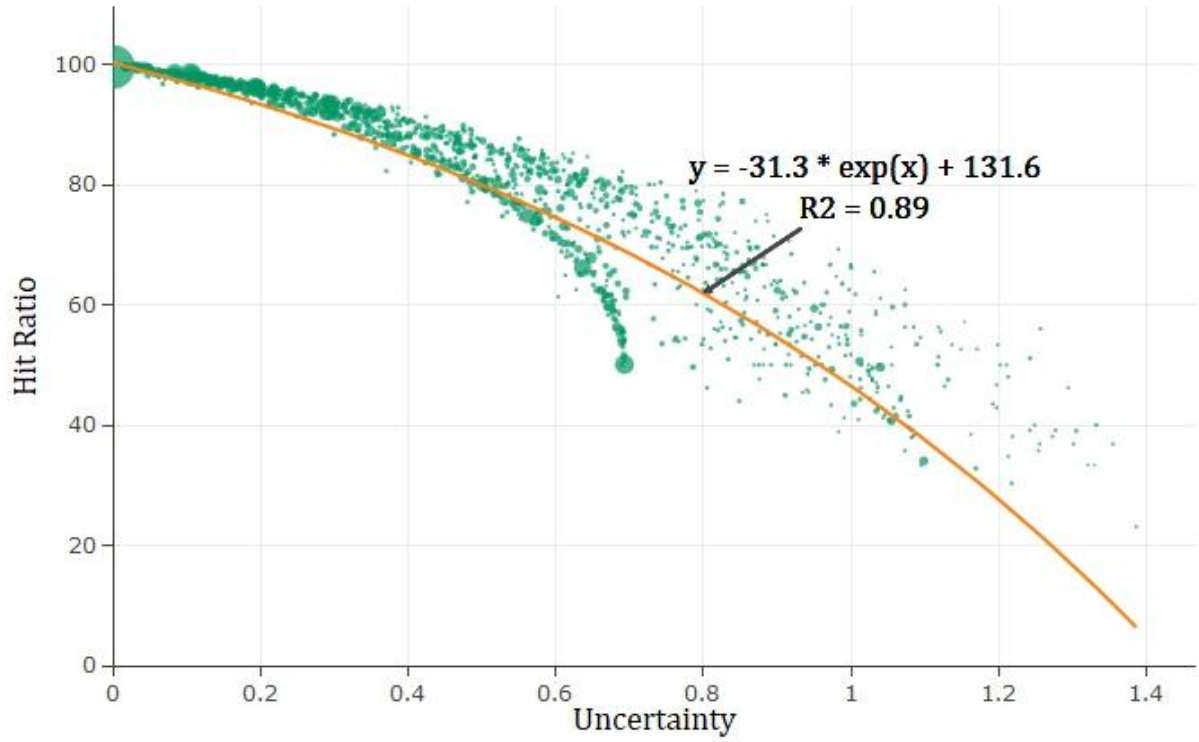


Figure 4.7. The hit ratio versus uncertainty value from every unique rule for DoTRules. Here, bubble size shows the frequency of rules with a same uncertainty value.

4.5. Discussion

4.5.1. DoTRules for LUCC simulation

The method used to calculate transition potential maps greatly affects the performance of automata-based LUCC models (Charif et al., 2012; Mas et al., 2014; van Vliet et al., 2016). Transparency and uncertainty of the transition rules are important for providing a better understanding of the model structure and of the nature of land-use transitions. Here, we introduced and applied a competent and transparent algorithm for the calculation of LUCC transition potential maps for use in cellular automata modelling of land-use change. As a predictive algorithm, DoTRules represents a new way to extract transition rules and map transition potential for use in LUCC simulation models while also enabling the mapping of uncertainty values at both pixel and rule levels. This makes uncertainty explicit and opens up the information contained in the transition rules to better model scrutiny and to better-informed land-use planning and policymaking. The potential of DoTRules is demonstrated here for the Ahvaz study area, and more applications and experiments are now required in different geographic contexts to fully explore the general applicability of DoTRules as a land-use transition rule detection

algorithm. Below we discuss expand on the advantages and limitations of using DoTRules for cellular automata-based LUCC simulation.

4.5.1.1. Transparency of transition rules

Depending on the method used to extract transition rules, some rule sets may be omitted which impairs the quality of the transition potential map and subsequent LUCC simulation. For instance, in applying black-box algorithms such as RF to calculate LUCC transition potential, it is difficult, if not impossible, to interpret the derived relationships between future land-use and predictor variables such as cell state, neighbourhood, and suitability variables. As a result, we cannot gain a clear understanding of the problem at hand due to the lack of an explanatory capability to provide insight into the characteristics of the target dataset (Qiu and Jensen, 2004). The aim of developing and applying DoTRules for LUCC simulation was to improve the quality of urban planning by identifying reliable and accessible land-use transition rules. DoTRules provides an opportunity to uncover model structure by adding more clarity to implemented transition rules and revealing information such as rules component priority order and values, frequency distribution of potential matching land-use labels for every rule, and rule-exclusive hit ratio and uncertainty (Figure 4.4). It can also identify simplified sub-rule sets compromising only the more informative variables. For instance, considering major types of land-use change/persistence, DoTRules identifies the rules governing the most common change/persistence patterns within the study area. Land-use planners and policy makers can explore alternative possibilities for land-use if past transparent and well-understood land-use transition rules continue into the future. A transparent set of land-use transition rules can also help uncover model structure hidden by other black-box algorithms and aid model verification and error identification (Bauer and Steinnocher, 2001; Tseng et al., 2008). Although most LUCC modellers are aware of the shortcomings and the limited accuracy of their input data, very little is known about the propagation of such errors in LUCC models (Dong et al., 2015).

4.5.1.2. Uncertainty of transition rules

We have demonstrated DoTRules' ability to quantify the frequency and uncertainty of land-use transition rules (Table 4.3; Figure 4.4). It is beneficial to gain reliable information corresponding to the functionality of those rules including their uncertainty and their corresponding hit ratio (Figure 4.4). Regardless of the fact that a rule is indicating the change or persistence of land-use in a grid cell, the rule-level uncertainty values provide a strong indication of prediction hit ratio and can be applied as a filter to remove low accuracy portions of a LUCC map derived from unreliable transitions.

The uncertainty of rules belonging to primary land-use transitions (Table 4.3) is also useful for the prior exploration of potential accuracy values (i.e. accuracy estimate) and can reduce or eliminate the need for post-hoc validation. This can be done for the primary rule set or a sub-rule set extracted from training data focusing on some specific land-use variables forming land-use transition potential maps. DoTRules' rule-level uncertainty product contains detailed information about the multiple specific land-use transitions in a study area, which provide landscape and urban planners with the opportunity to better understand the nature of future land-use transitions and the level of confidence in their prediction. For instance, considering the results of uncertainty assessment for LUCC simulation (Figure 4.6 and Table 4.3), the land-use transition between urban and agriculture/bare lands is more uncertain than simulated land-use transitions between agricultural lands and bare lands. By specifying uncertainty thresholds, land-use transitions in which we are most confident or least confident can be identified, and these can be made explicit in mapped planning outputs to guide decision-making. Thereby, uncertainty thresholds may be applied to any type of transitions for reducing the rule population to include only the most trusted (i.e. those below a specific uncertainty value). The high degree of correspondence between uncertainty and hit ratio means that planners can use DoTRules' uncertainty products as an indicator of land-use simulation accuracy. For example, if the uncertainty threshold of 0.2 is applied, then planners can be confident that the hit ratio of all the remaining rules should be high (i.e. above 90%) (Figure 4.7).

4.5.1.3. Mapping uncertainty

In addressing the limitations of widely used accuracy measures and indices, using the DoTRules algorithm, we can also map and apply localised uncertainty values within different classification stages as an estimate of accuracy. Thus, a strength of DoTRules in LUCC modelling is its demonstrated ability to quantify the uncertainty of simulated land-use patterns at pixel level, regardless of test data availability. Uncertainty maps can be produced as an estimate of prediction accuracy even for future land-use change scenarios which lack ground truth data. This enables the most recent data to be used in model-building rather than model validation, which should produce more reliable land-use simulation further into the future, yet still providing an estimate of uncertainty. Spatially-explicit uncertainty also assists landscape and urban planners to foresee the degree of susceptibility to prediction error for specific localities. Rule-level uncertainty maps are created for every LUCC iteration (30 iterations in the present study), as every pixel corresponds to one transition rule for every iteration. The mean uncertainty of these transition rules allocated to a single grid cell may be calculated, mapped and analysed for one or many iterations. Uncertainty values vary over time for LUCC simulation for each pixel and this can be graphed over time. This helps landscape and urban planners to keep track of LUCC modelling uncertainty and/or hit ratio at pixel level at different time steps of a LUCC model. Other indicators of uncertainty may also be produced including the maximum, minimum, or median; each of which may be useful, depending on the planning application.

4.5.2. Broader application to environmental modelling

DoTRules quantifies the level of correspondence between each predictor variable and response variable (Table 4.) through the calculation of entropy values (eq. 4.3). In addition, it also extracts the transparent rules, each with a quantified frequency and entropy, which provides insight into the observed dynamics (Table 4.3). In this paper, we have applied DoTRules for calculating transition potential in cellular automata models of land-use/cover change. Nonetheless, the conceptual framework of DoTRules can be applied to other applications of automata models such as plant population dynamics, forest/bush fire spreading, slope failure, debris flow, and urban sprawl where transition potential mapping is required and where it is necessary to understand the direction and magnitude of state transitions with greater transparency. For instance, in terms of bushfire simulation using cellular automata, a same set of transition rules should be determined to form a transition potential map (Quartieri et al., 2010) while different set of predictor variables such as bush density, flammability, land height and wind (Li and Magill, 2001). Simultaneously, the application of rule-base methods -such as DoTRules- for a better understanding of dynamic environmental phenomena is not limited to automata models (Amini et al., 2010; Luo et al., 2017b; Moore et al., 2014) and it involves a broader range of environmental models. The authors are currently trialling DoTRules in other critical applications such as image classification and landslide susceptibility zonation.

4.5.3. Limitations of DoTRules

The major limitation of DoTRules corresponds with the calculated entropy value for low-frequency rules where, in calculating Shannon entropy values, some rules may be attributed a low-uncertainty value by chance, due to a very low-frequency value. For instance, if the entropy value of two different rules is equal, they would be attributed the same uncertainty class. Nonetheless, they may be different in terms of frequency values (Figure 4.8). In this regard, among the rules with the same entropy values, those with higher frequency would be more reliable as they occurred more often. A subjective thresholding approach may be beneficial to deal with this problem. Thus, those low-frequency rules should be ignored if they belong to a frequency value lower than a threshold defined by the user.

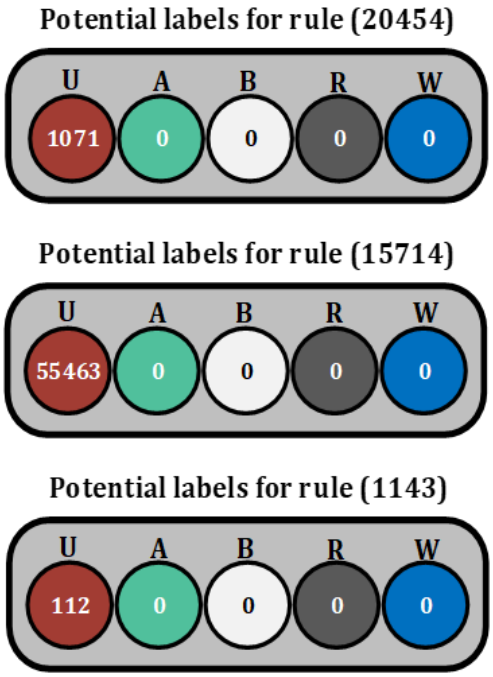


Figure 4.8. Different rules with the same uncertainty measure (entropy value = 0) and different frequency values. Here, letters above the circles represent land-use/cover types such as U): urban, A): agriculture, B), bare lands, R): roads and W): water.

4.6. Conclusion

Cellular automata have long been used to capture the complex dynamics of LUCC processes. We have presented a new and innovative algorithm called DoTRules—Dictionary of Trusted Rules—for calculation of transition rules and transition potential maps for use in CA-based simulation models. We have then presented an application of the proposed approach in the context of LUCC simulation where cellular automata models are increasingly popular. DoTRules enables land-use allocation to be implemented with a new level of transparency and transition rule characteristics are accessible and can be monitored. DoTRules also enables the spatial exploration of LUCC prediction uncertainty. These estimates can assist urban planners in avoiding risky land-use predictions where rules are not reliable. Assessing rule information also enables the detection of trends and understanding processes of land-use change in a study area. The performance of automata simulation based on DoTRules transition potential calculation was very similar to simulation based on the state-of-the-art random forest method. Hence, we conclude that DoTRules is a promising approach for extraction of transition rules and providing transition potential maps for CA-based simulation due to its predictive ability, transparency, and ability to produce multiple uncertainty products. These capabilities can enhance the utility of automata-based simulation as a tool for end-users and improve the quality of the resultant decision-making process. More experiments in different environmental applications with a different set of variables are now required to verify the broader utility of DoTRules as a CA calibration tool. The application of DoTRules as a transparent rule-based approach to other environmental modelling applications should also be explored.

5. Drought sensitivity mapping using two one-class support vector machine algorithms

Chapter 5 applies a novel drought sensitivity mapping method using two one-class support vector machine algorithms where fuzzy If-Then rules are benefitted to combine resultant outcome of each one-class support vector machine (OC-SVM). This chapter is an attempt to achieve objective number four through modelling uncertainty Type B. The current chapter has been published in **Atmospheric Research** journal 17th May 2017.

Roodposhti, M.S., Safarrad, T. and Shahabi, H., 2017. Drought sensitivity mapping using two one-class support vector machine algorithms. Atmospheric Research, 193: 73-82.

Abstract: This paper investigates the use of the standardised precipitation index (SPI) and enhanced vegetation index (EVI) as an indicator of soil moisture. On the other hands, we try to produce a drought sensitivity map (DSM) for vegetation cover using two one-class support vector machine algorithms. In order to achieve promising results, we have used a combination of both 30 years statistical data (1978 to 2008) of synoptic stations and 10 years MODIS imagery archive (2001 to 2010) within the boundary of Kermanshah province, Iran. The synoptic data and MODIS imagery were used for extraction of SPI and EVI, respectively. The objective is therefore to explore meaningful changes of enhanced vegetation index (EVI) in response to drought anomalies, in the first step, and further extraction of reliable spatio-temporal patterns of drought sensitivity using efficient classification technique and spatial criteria, in the next step. To this end, four main criteria including elevation, slope, aspect and geomorphic classes are considered for DSM using two OC-SVM algorithms. Results of the analysis showed distinct spatio-temporal patterns of drought impacts on vegetation cover. The receiver operating characteristics (ROC) curves for the proposed DSM was used along with the simple overlay technique for accuracy assessment phase and the area under curve (AUC =0.80) value was calculated.

5.1. Introduction

Global warming affects evapotranspiration, which is the movement of water into the atmosphere from land, water surfaces and plants due to evaporation and transpiration. This is expected to increase to both drought severity measure and geographic expansion of dry areas. When discussing drought, one must have a proper understanding of aridity and the difference between the two. Climatologically, aridity is defined as "the degree to which a climate lacks effective, life-promoting moisture" while drought is "a period of abnormally dry weather sufficiently long enough to cause a serious hydrological imbalance" (Hayes et al., 2011). Aridity is measured by comparing long-term average precipitation and evapotranspiration. It is obviously a permanent climatic characteristic. In this regard, the arid climate indicates that average long-term evapotranspiration is greater than average long-term precipitation value. On the other hands, drought refers to the moisture balance that is mainly estimated on the annual, seasonal or monthly basis. As opposed to aridity, drought is a transient climatic idiosyncrasy (Lioubimtseva and Adams, 2004). Despite the apparent simplicity of this definition, due to its long-term development and duration, the progressive characteristics of its impacts and spatial extent, drought is the most complex natural hazard to identify, analyse, monitor and manage (Burton, 1993; Vicente-Serrano et al., 2012; Wilhite, 2012).

Nowadays, with progressive human development and subsequent climate change, drought monitoring and impact assessment program is of great importance. In order to reduce the drought vulnerability of the affected regions, it is vital to truly comprehend spatio-temporal drought patterns and their subsequent impacts. This will facilitate fulfilments of further measures focused on promoting both drought risk mitigation and preparedness. Risk mitigation simply refers to long-term measures for reducing the risk including the development of technological solutions, legislation, land-use planning, insurance, etc. (Vicente-Serrano et al., 2012). Basically, risk mitigation measures should be

implemented through accurate identification of risks and promotion of the risk perception which are acquired within preparedness phase (Bird, 2009). Preparedness refers to the development of various emergency plans and warning systems aimed for efficient decision making and acting once the disaster strikes or even it is anticipated.

During the last decades, different types of drought indices (DIs), including climatic or satellite-derived DIs for regional to global scale drought assessment and monitoring have been developed and implemented. Climatic DIs such as percent of normal, standardized precipitation index (SPI) (Cancelliere et al., 2007; Cui et al., 2015; Sönmez et al., 2005; Zhang et al., 2009), deciles, palmer drought severity index (PDSI) (Dai, 2011; Dai et al., 2004; Li et al., 2007), crop moisture index (CMI) (Keyantash and Dracup, 2002; Quesney et al., 2000), surface water supply index (SWSI) (Shafer and Dezman, 1982; Toulios et al., 2012) etc. aimed to unify thousands of bits of climatological data such as rainfall, snowpack, stream flow and other water supply indicators into a comprehensible quantitative measure. A climatic DIs typically depicts how much the climate of specified geographic location in a given period has deviated from historically established normal conditions (Jairath, 2008; Mu et al., 2013; Narasimhan and Srinivasan, 2005; Pai et al., 2011; Werick et al., 1994). On the other hand, variety of satellite DIs including: normalized difference vegetation index (NDVI), vegetation condition index (VCI), temperature condition index (TCI), enhanced vegetation index (EVI) etc. also developed to quantify the drought impact on soil and/or natural vegetation cover (Bhuiyan et al., 2006; Chang et al., 2010; Huete et al., 2002; Toulios et al., 2012).

Introduction and evaluation of novel spatial methodologies for the identifying the measure of persistence and resilience of an ecosystem despite climate change constitutes a research priority of global relevance. As drought is identified one major issues of global climate variability, therefore, we present a novel approach to assess the relative sensitivity of vegetation cover, as a major constituent of any ecosystems, to drought.

Since many different DIs have been used in drought monitoring and drought sensitivity mapping studies, in an attempt to find an optimal solution and to consolidate the accuracy of obtained results, we calculated and further compared two different DIs including: 1. The standardised precipitation index (SPI) as a climatic DI and 2. Enhanced vegetation index (EVI) as satellite-derived DI. Afterwards, we made an effort to assess vegetation cover sensitivity to drought which is simply ecosystem sensitivity to short-term climate variability and regions of amplified vegetation response (Seddon et al., 2016). A novel method to identify different sensitivity classes of a drought sensitivity map (DSM) with respect to changes in elevation, slope, aspect and geomorphic criteria is implemented for this purpose. Two one-class support vector machine (OC-SVM) algorithms were beneficially used to obtained the final DSM.

The paper is organized as follows: after a description of the study area in Section 5.2, a detailed definition of the material and methods of the research is described in Section 5.3. Section 5.4 presents results while Section 5.5 belongs to a short discussion and conclusions.

5.2. Description of study region

Kermanshah province is located between $32^{\circ}36'$ to $35^{\circ}15'$ N latitude and $45^{\circ}24'$ to $48^{\circ}30'$ E longitude in north-western Iran and it is considered a part of the structural zone of Folded Zagros in the boundary of Arabian and Iran plate (Fig. 5.1). It has a climate which is heavily influenced by the proximity of the Zagros Mountains, classified as a hot dry summer with rather cold winters and there are usually rainfalls in fall and spring. Kermanshah climate is classified as typical Mediterranean climate (Csa) in Köppen-Geiger classification.

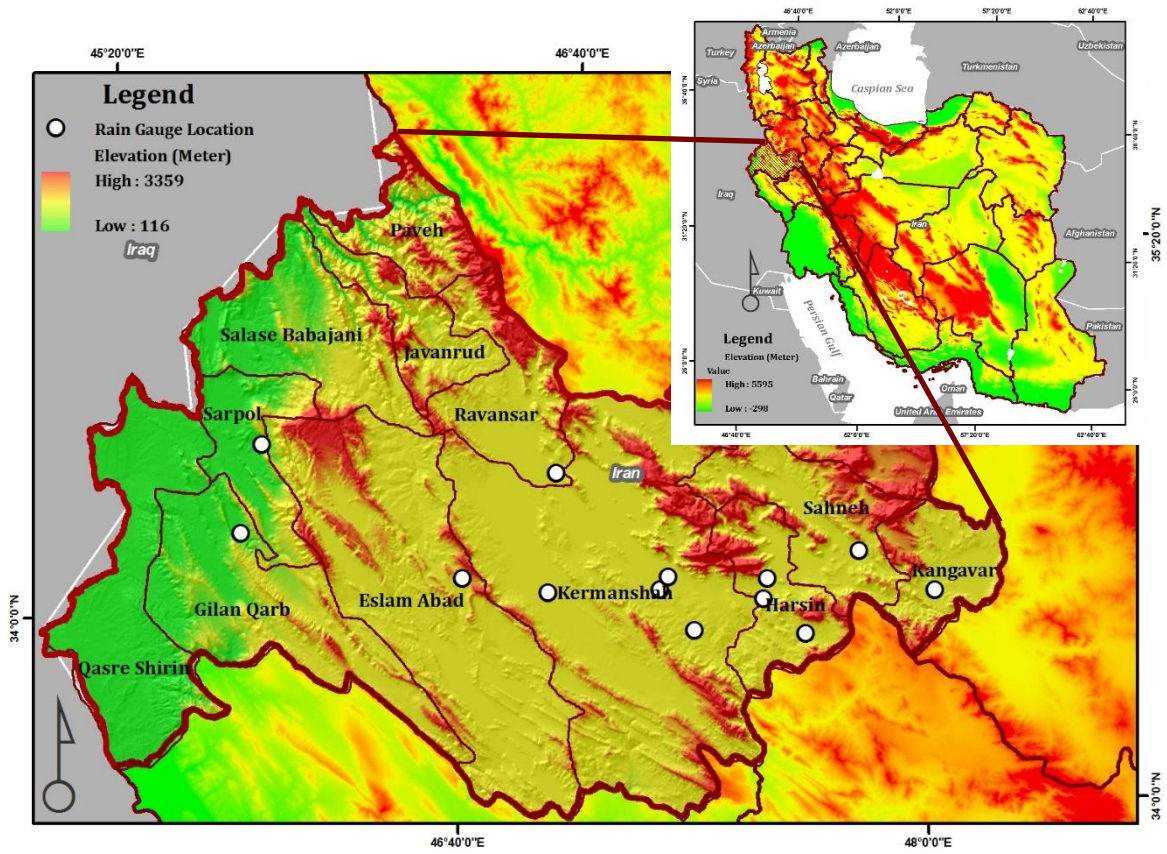


Figure 5.1. Location map of the study area

The minimum amount of precipitation occurs in June with the average of 0 mm while the maximum amount occurs in March, with an average of 76 mm. The city's altitude and exposed location relative to westerly winds makes precipitation a little bit high with a total annual precipitation average 478.7 mm. However, at the same time, it produces huge diurnal temperature swings especially in the virtually rainless summers, which remain extremely hot during the day. The annual average temperature of Kermanshah is 14.2°C while the average monthly temperatures vary by 27.6 °C. Kermanshah also experiences snow cover for at least a couple of weeks in winter.

Arid and semi-arid regions cover almost 40 % of the world's land (Aydin, 1995; Bannayan et al., 2010) and according to the climate models during the twenty-first century in the semi-arid Mediterranean, severe water loss will be caused by the climate variability (Houghton et al., 2001). Therefore, Kermanshah province of Iran could be an ideal study area for drought studies.

5.3. Material and methods

5.3.1. Data

Here, posterior to the extensive review of the relevant literature; we selected three main categories of related geospatial data. First, 30 years (from 1978 to 2008) collection of mean monthly precipitation obtained from 13 different Iranian Meteorological Organization (IMO) stations (i.e. climatological, synoptic and rain gauge stations) throughout Kermanshah province. Second, MOD 13 or enhanced vegetation index (EVI) of Terra series to effectively use as a satellite-derived DI for 10 years from 2001 to 2010. Third, 90-meter resolution digital elevation model (DEM) products of the Shuttle Radar Topography Mission (SRTM) and geomorphic unit maps of the study region used for extraction of spatio-temporal changes of drought patterns.

We also used an inventory of the most negative and positive changes in EVI value containing 842 first negative along with 842 first positive points of EVI change for both training OC-SVM algorithm and further validation of proposed DSM. It also must be noted that database of the most negative and positive changes in EVI was produced by differentiating EVI values of mild and severe drought which have accrued during the 2006 and 2008. In terms of mentioned EVI inventory, it was randomly split into a train (≈ 67%) and test (≈ 33%) samples for training the proposed DSM and subsequent validation purpose, prior and posterior map elaboration, respectively.

5.3.2. Methodology

The proposed methodology has three steps: in step 1, we calculated the standardised precipitation index (SPI), as one of the most widely used climatic DIs. Accordingly, in step 2, to establish the satellite-derived DI of considered study region, the EVI was used. Finally, the obtained results of the two first steps were used for subsequent analysis of spatio-temporal changes of drought impact on vegetation cover and a drought sensitivity map was then produced using two one-class support vector machine algorithms in step 3 (See Fig. 5.2).

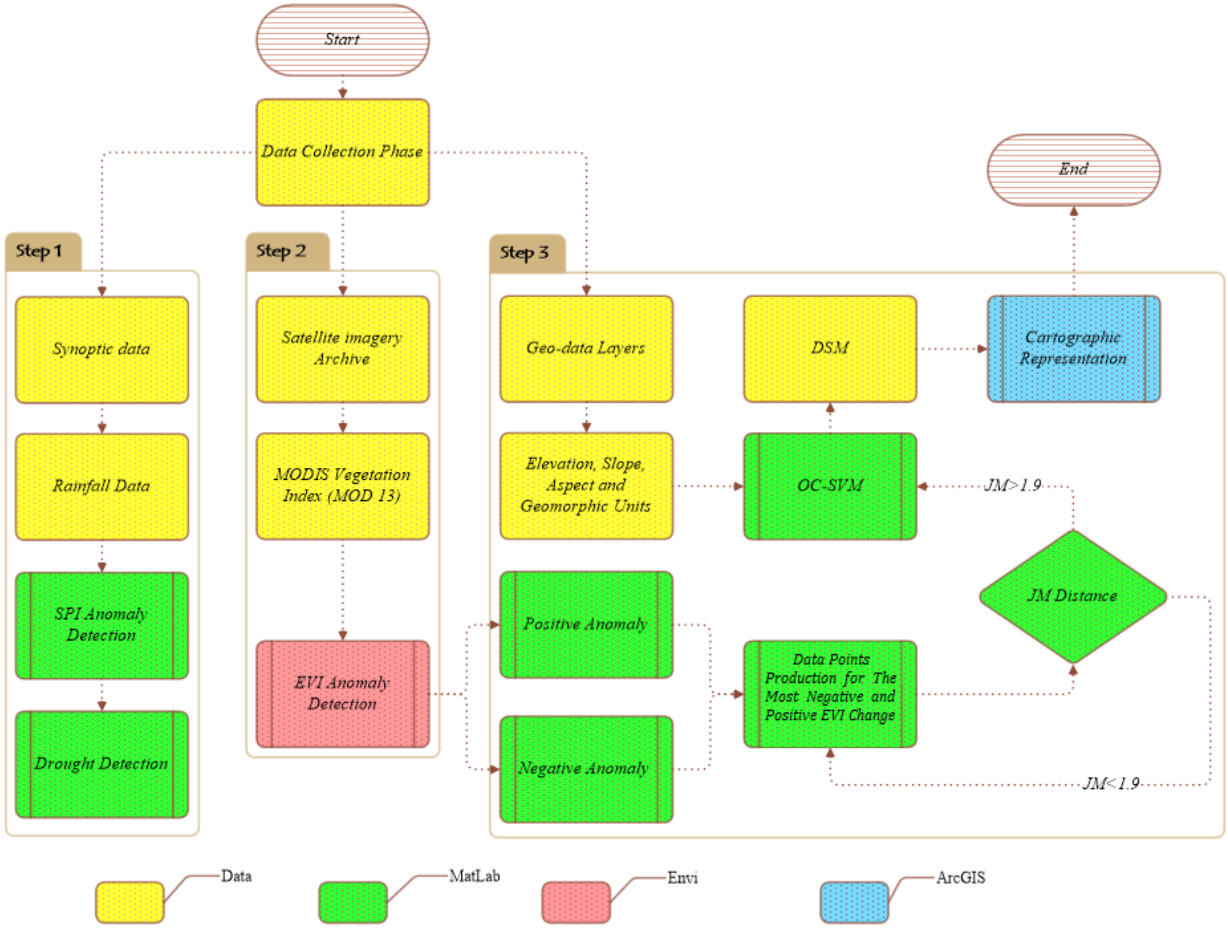


Fig. 5.2. Schematic representation of the 3-step methodology implementation.

5.3.2.1. Standardised Precipitation Index (SPI)

The standardised precipitation index (SPI) was proposed by (McKee et al., 1993) and if it is not the most widely used DIs (Belal et al., 2014; Jain et al., 2010; Karavitis et al., 2011), it is considered as one of most popular DIs of drought monitoring and assessment (Jiang et al., 2008; Liu et al., 2013; Mu et al., 2013; Vicente-Serrano et al., 2012). SPI is based on the probability of precipitation for any time scale. In

other words, it is a simple index which is the number of standard deviations that the observed precipitation deviates from the long-term mean (i.e. 30 years), assuming a normal distribution. The SPI was designed for quantification of precipitation deficit for various time scales. These time scales reflect the impact of drought on the availability of the different water resources. Soil moisture conditions respond to precipitation anomalies on a relatively short scale. Groundwater, stream flow, and reservoir storage reflect the longer-term precipitation anomalies. Accordingly, McKee et al. (1993) originally calculated the SPI for 3, 6, 12, 24, and 48–month time scales (See Table 5.1).

Table 5.1 Drought category of SPI value (McKee et al., 1993).

SPI value	Drought category	Time in category (%)
0 to −0.99	Mild	34.1
−1.00 to −1.49	Moderate	9.2
−1.50 to −1.99	Severe	4.4
≤ −2.00	Extreme	2.3

5.3.2.2. Enhanced Vegetation Index (EVI)

The enhanced vegetation index (EVI) is one of the most popular satellite-based vegetation indices produced for the Terra and Aqua Moderate Resolution Imaging Spectroradiometers (MODIS). It is an 'optimised' index designed to enhance the vegetation signal with improved sensitivity in high biomass regions. It further improves vegetation monitoring through minimising both soil and atmosphere influences (Jiang et al., 2008). EVI is one of widely used satellite vegetation greenness indices which have been successfully used to monitor global vegetation photosynthetic activity (Huete et al., 2002; Jiang et al., 2008; Justice et al., 2002; Tucker, 1979). Drought indices derived from satellite imagery have been widely used to identify spatial extents of drought (Vergni and Todisco, 2011). The indices are useful for detection and monitoring large area vegetation stress resulted from drought or soil oversaturation following flooding and excessive rains. This soil and atmosphere resistant vegetation index is defined by (Solano et al., 2010):

$$EVI = G\left(\frac{P_{nir} - P_{red}}{L + P_{nir} - C_1 P_{red} - C_2 P_{blue}}\right) \quad (5.1)$$

where P is 'apparent' (top-of-the-atmosphere) or 'surface' directional reflectance, L is a canopy background adjustment term equal to 1, G gain factor and another constant equal to 0.2 and C1 and C2 weigh the use of the blue channel in aerosol correction of the red channel estimated about 6 and 7.5, respectively.

5.3.3. One-class support vector machine

Support vector machine (SVM) is a supervised learning method derived from statistical learning theory and the structural risk minimization principle (Boser et al., 1992; Vapnik, 2013; Vapnik and Vapnik, 1998). It uses a decision surface to separates the target classes through maximisation of the margin between them (Burges, 1998). The mentioned surface is usually called the optimal hyperplane, and the data points closest to the hyperplane are called support vectors (Fig. 5.3). It should be noted that the support vectors are considered critical elements of the training set.

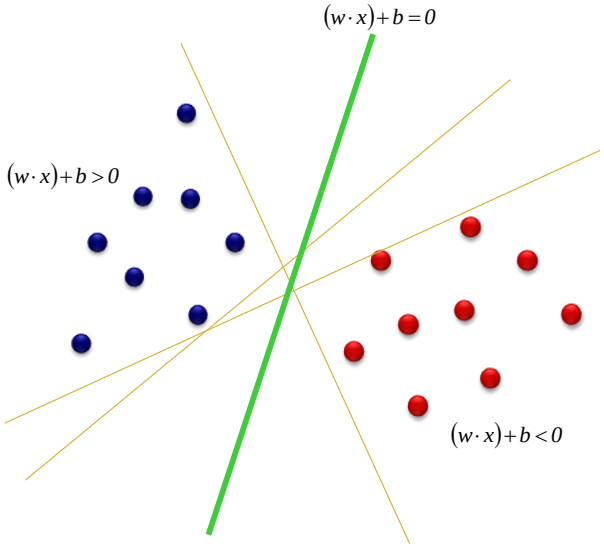


Fig. 5.3. Optimal separating hyperplane (Gunn, 1998).

However, often support vector machines are an example of a linear two-class algorithms are aimed to maximize the margin between the two classes (Fig. 5.4a), it could be used for one-class classification purpose, where one tries to detect one class and reject the others (Fig. 5.4b) (Gunn, 1998; Muñoz-Marí et al., 2010).

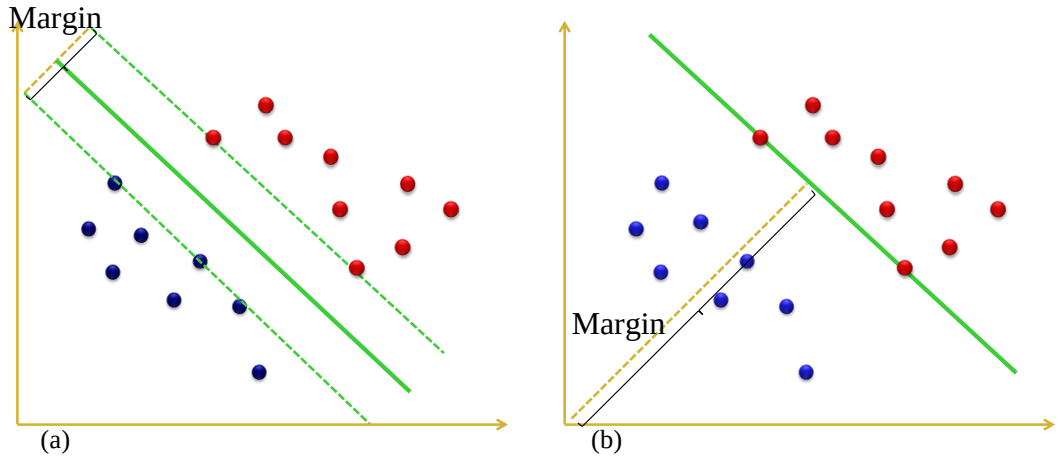


Fig. 5.4. Typical support vector machines classifiers: a) Two Class SVM (Gunn, 1998) b) One Class SVM (Muñoz-Marí et al., 2010).

The one-class support vector machine (OC-SVM) was proposed by (Schölkopf et al., 2001) to estimate a set that encloses most of a given random sample where $x_i \in R^d$. Each x_i is first transformed via a map $\phi : R^d \to H$ where *H* is a high (possibly infinite) dimensional Hilbert space generated by a positive-definite kernel $k(x_i, y_i)$. The kernel function corresponds to an inner product in *H* through $k(x_i, y_i) = \langle \phi(x_i), \phi(y_i) \rangle$ .The OC-SVM speculates a hyperplane in the feature space which detaches the data from the origin with maximum possible margin (Figure 4b). In the event that no such hyperplane exists, slack variables ξ_i allow for some points to be within the margin, and the free parameter $\nu \in [0,1]$ controls the cost of such violations. In fact, ν can be shown to be an upper bound on the fraction of points within the margin (outliers) (Schölkopf et al., 2001). The hyperplane in feature space induces a generally nonlinear surface in the input space. More precisely, the OC-SVM as presented in

(Schölkopf et al., 2001) (Tax and Duin, 1999) requires the solution of the following optimisation problem:

$$\min_{w,\xi,b}\left\{\frac{1}{2}\|w\|^2+\frac{1}{vl}\sum_{i=1}^{1}\xi_i-p\right\} \tag{5.2}$$

$$Subject-to\ \ (w,\Phi(x_i))\geq p-\xi_i, \tag{5.3}$$

$$\xi_i\geq 0.$$

Here, ω is a vector perpendicular to the hyperplane in H, and ϱ is the distance to the origin. Since the training data distribution may contain outliers, a set of slack variables $\xi_i \geq 0$ is introduced to deal with them (which allows for penalised constraint violation), as usual in the SVM framework. The parameter $v \in [0,1]$ controls the trade-off between the number of examples of the training set mapped as positive by the decision function $f(x)=sng(w,\Phi(x_i))-p$ and having a small value of $\|w\|$ to control model complexity. Finally, it must be noted that it is possible to segregate two patterns either through one two-class support vector machine (TC-SVM) or two OC-SVMs, which the latter produces more conservative decision regions (Elshinawy et al., 2010).

5.3.4. Kernel Functions

The performance of the SVM model depends on the choice of the kernel parameters. Accordingly, selection of the kernel function is very important in SVM modelling (Xu et al., 2012). However new kernels are being proposed by researchers, four kinds of them are often used: linear kernel, polynomial kernel, RBF kernel (often called Gaussian kernel) and sigmoid kernel as the last one (Brereton and Lloyd, 2010; Hsu et al., 2003). In the present study, we have chosen RBF kernel as the most popular kernel functions of the SVM algorithm.

5.3.4.1. Radial basis functions (RBF)

RBF has received significant attention in various kernelized learning algorithms (Hsu et al., 2003). It is simply defined as:

$$k(x_i,x_j)=exp\left(-\gamma\left\|x_i,x_j\right\|^2\right),\gamma\succ 0 \tag{5.4}$$

Where $\gamma \succ 0$ is a parameter that controls the width of Gaussian distribution. It plays a similar role as the degree of the polynomial kernel in controlling the flexibility of the resulting algorithm (Ben-Hur and Weston, 2010).

5.3.5. Jeffries-Matusita (JM) separability measure

With respect to the fact that drought sensitivity mapping was the goal of the present study, the assessment focused on distinguishing the most sensitive from the least sensitive class. Here, separability degree of sensitivity classes was assessed through Jeffries-Matusita (JM) separability approach that used both training subsets including more and less sensitive location. The JM distance between a pair of class-specific probability functions is defined as following (Richards and Jia, 1999a):

$$J_{ij}=\int_x\left(\sqrt{p(x|w_i)}-\sqrt{p(x|w_j)}\right)^2 dx \tag{5.5}$$

where $p(x|w_i)$ and $p(x|w_j)$ are conditional probability density functions for the feature vector x, given in data classes of w_i = more sensitive and w_j = less sensitive events respectively. Under normally distributed classes this becomes:

$$J_{ij} = 2\left(1 - e^{-B}\right) \tag{5.6}$$

where

$$B = \frac{1}{8}\left(m_i - m_j\right)^t \left\{\frac{\Sigma i + \Sigma j}{2}\right\}^{-1} \left(m_i - m_j\right)$$

$$+ \frac{1}{2} In\left\{\frac{\left|\left(\Sigma_i + \Sigma_j\right)/2\right|}{\left|\Sigma i\right|^{\frac{1}{2}} + \left|\Sigma_j\right|^{\frac{1}{2}}}\right\} \tag{5.7}$$

In this notation, m_i and m_j correspond to class-specific, expected sensitivity values, and $\sum_i$ and $\sum_j$ are unbiased estimates for the class-specific covariance matrices of more and less sensitive subsets respectively, *ln* is the natural logarithm function, $\left|\sum i\right|$ and $\left|\sum_j\right|$ are the determinant of $\sum_i$ and $\sum_j$ (matrix algebra). JM separability measure takes on a maximum value of 2.0, and values above 1.9 indicate assuring separability (Richards and Jia, 1999a). For lower separability values, it should be taken into consideration to improve the separability by editing the position of more sensitive points which are located in less sensitive areas, vice versa. Here, the achieve value for the more and less sensitive separability measure was equal to 1.976 which suggests that the two more and less sensitive training subsets may be distinct with high separability.

5.4. Results

When drought occurs through an extended period, a region receives a deficiency in its water supply, whether atmospheric, surface or ground water. In other words, as a result of drought stream and river flows decline, water levels in lakes and reservoirs fall, and the depth to water in wells increases. This will further leads to decrease in soil moisture, which in long-term primary controls vegetation and ecosystems. Unlike the above mentioned immediate impacts of drought, however, long-term impacts could be harder to monitor and more costly to manage in the future. Here, we have used 30-years of monthly precipitation data for SPI calculation which was used to quantify the precipitation deficits for multiple time scales.

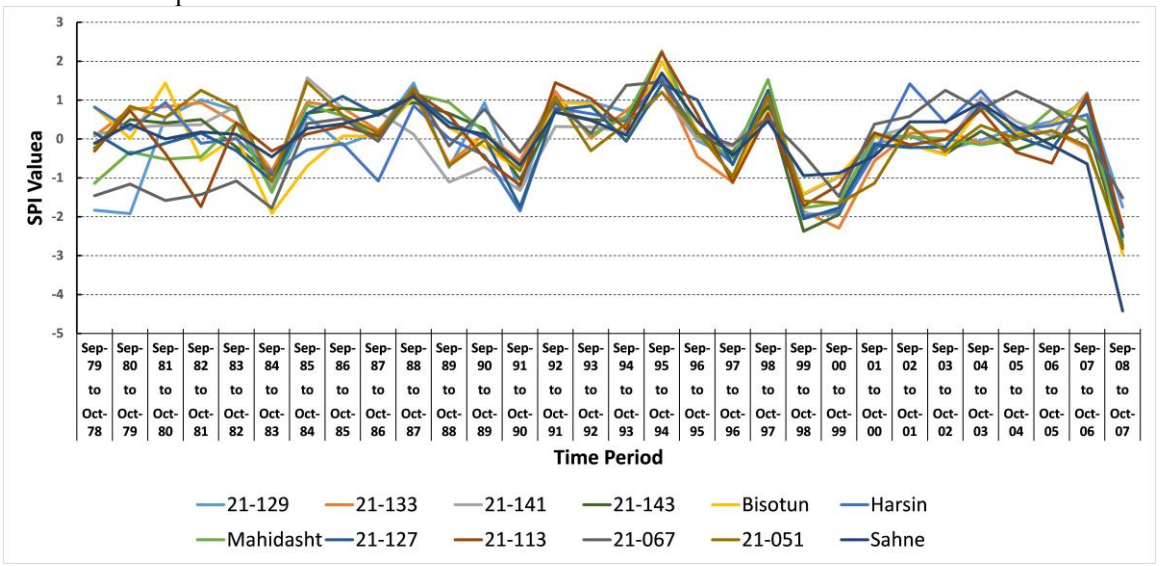


Fig. 5.5. Standardised precipitation index (SPI) for October 1978 through September 2008 with a time scale of 1 months.

In this study, the SPI has been calculated for time-scales of 1 month for each year. Figure 5.5 shows the SPI values for a time-scale of 1978 through 2008. As can been seen from Figure 5.5, the minimum SPI value indicates a severe drought occurred in 2008 while the maximum SPI value belongs to 1995. Afterwards, we have also examined the spatio-temporal patterns of EVI values of late-April (the most monthly EVI value) and mid-October (the least monthly EVI value) through the entire study region from 2001 to 2010. Results of both SPI and EVI testify a severe drought occurrence in 2008 which was selected as the basis of DSM (Fig. 5.6).

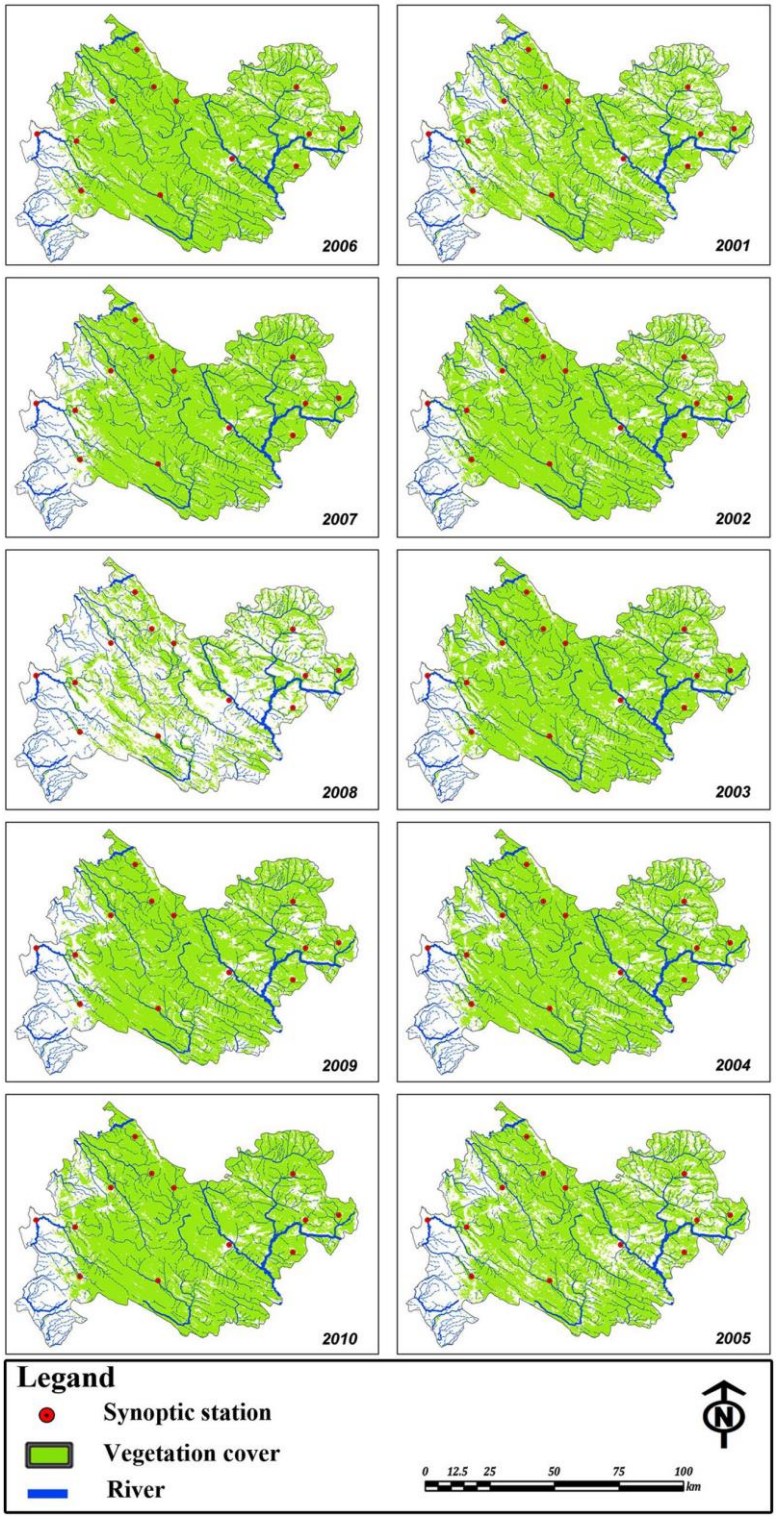


Fig. 5.6. Annual EVI values of late-April from 2001 to 2010.

In other words, considering the fact that there is no satellite imagery archive belonging to 1995 (i.e. maximum SPI value), we tried to use a pixel based ratio of average and minimum EVI values (i.e. 2006 and 2008) for further identification of the least and the most vegetation loss locations. As EVI is found to be linearly correlated with green leaf area index (LAI) in crop fields (Jiang et al., 2008), it will be maximized when it is representing highest leaf biomass. Therefore, here in the present study, pixels with most positive changes in EVI value, representing an increase in the biomass amount of the pixel, were detected as the less sensitive pixels to drought. However, pixels with the most negative changes were considered as more sensitive (Fig. 5.7). In other words, considering the fact that leaf biomass will be reduced as results of drought impacts, the word 'less sensitive' has been chosen when leaf biomass is not negatively affected by drought. Nonetheless, the word 'more sensitive' is selected for the opposite condition.

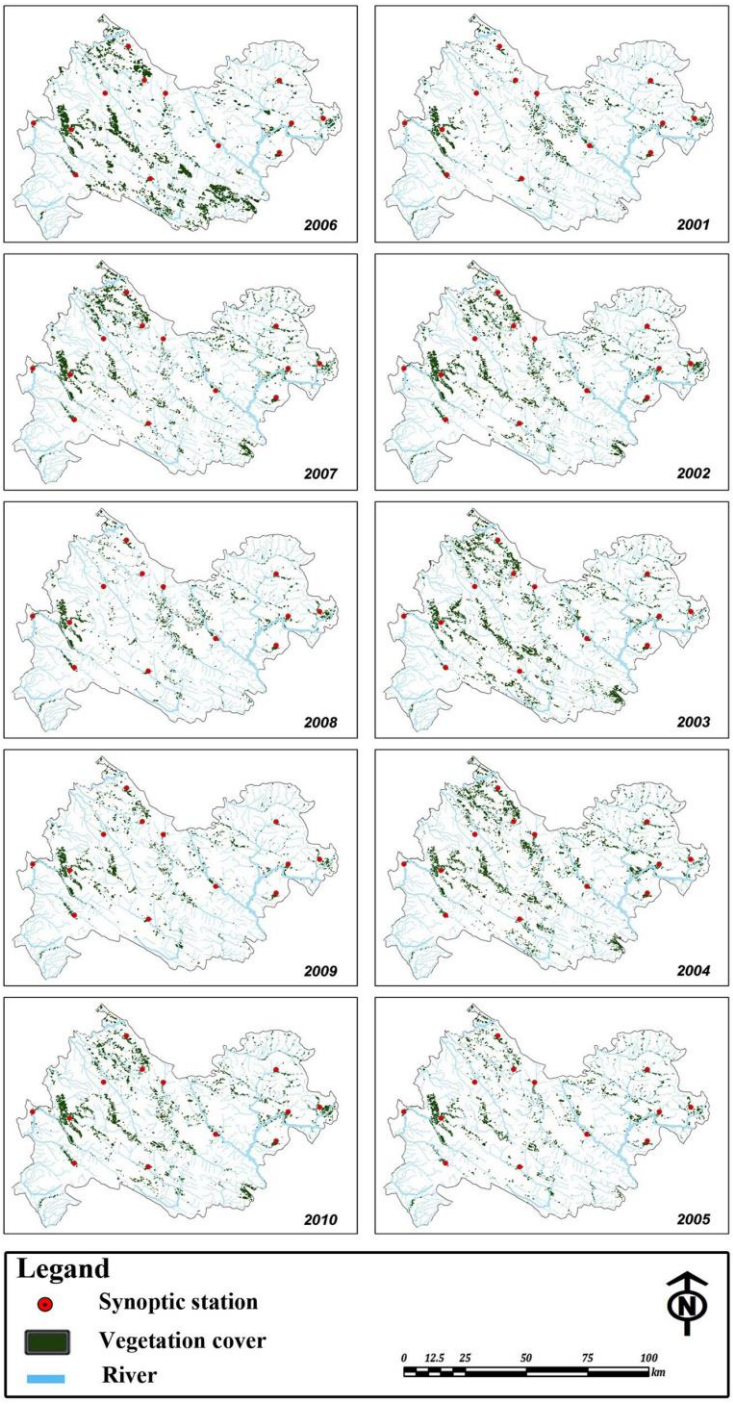
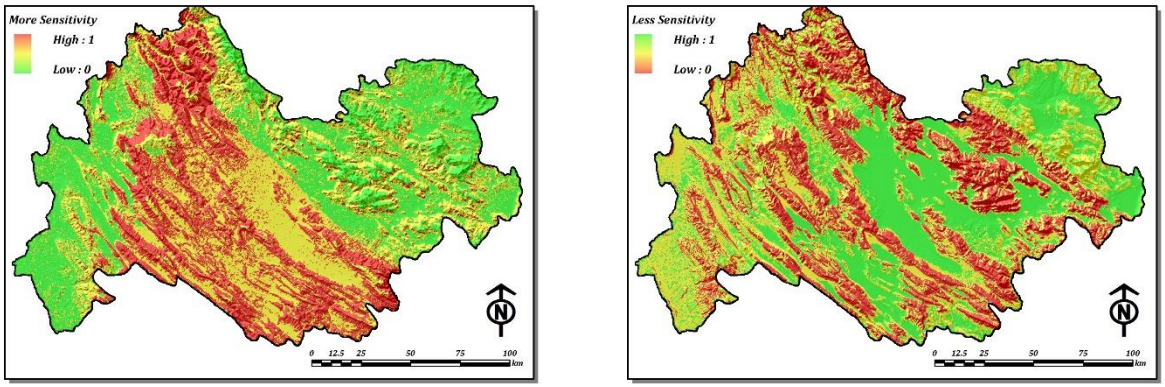


Fig. 5.7. Annual EVI values of mid-October from 2001 to 2010.

The proposed model of DSM is based on the OC-SVM classification technique that could be considered as a quantitative soft computing method within which less subjectivity is guaranteed. In this respect, following the accomplishments of necessary data pre-processing steps, each criterion of the study area was divided into a 278m*278m square grid, which contains 320,563 pixels, laid out in an irregular boundary. Accordingly, after importing the pre-processed data into the MatLab environment, an evaluation matrix is then constructed to be used in the classification process. Experimental results not only showed that OC-SVM is more efficient compared to TC-SVM algorithm while producing results of similar accuracy, but also it requires less time and storage space to run compared to TC-SVM (Manevitz and Yousef, 2002; Senf et al., 2006). As a result, two OC-SVMs were applied by an RBF kernel function to construct the respective DSM in a further step (Fig. 5.8).



(a) (b)

Fig. 5.8. Resultant outcome of each OC-SVM including: (a) more, and (b) less sensitive classes of proposed DSM scheme.

Finally, in the present study, the proposed DSM is divided into five sensitivity classes namely very low, low, moderate, high and very high using 2D scatter plots and nine fuzzy if-then rules (Fig. 5.9).

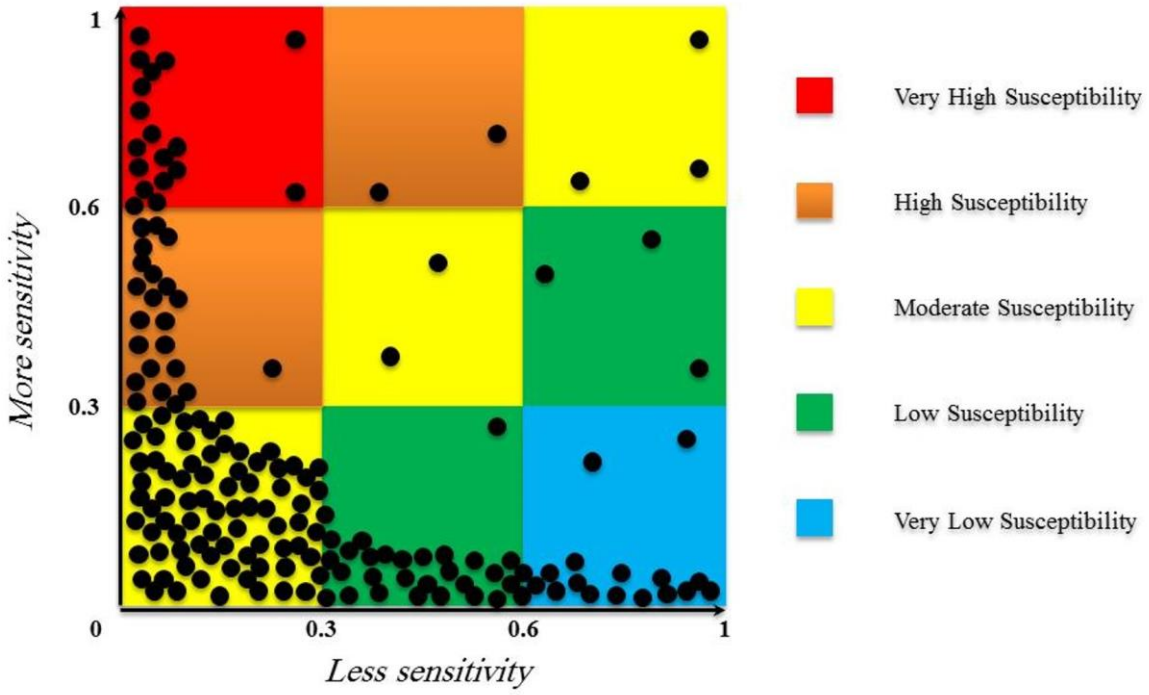


Fig. 5.9. Selected if-then rules for 2D scatter plots to interactively classify two categories of more and less sensitive pixels.

Figure 5.9 illustrates how fuzzy if-then rules have been used for pattern classification problems while Figure 5.10 represents final DSM of Kermanshah Province which has been elaborated using OC-SVM algorithms and corresponding geo-data layers including EVI, elevation, slope, aspect and geomorphic units.

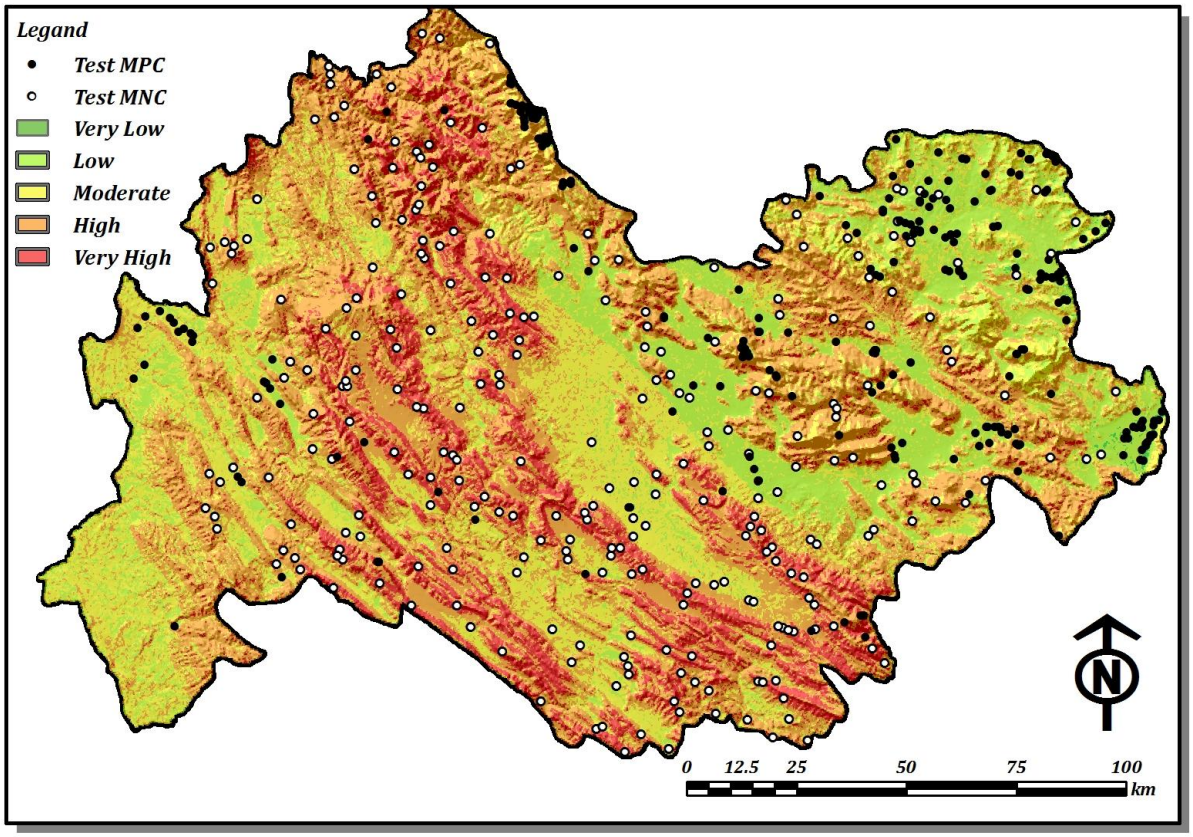


Fig. 5.10. Resultant sensitivity map of proposed DSM scheme along with the most negative and the most positive changes test data points of EVI.

5.4.1. Validation of the results

Validation is truly an essential step in the development of any predictive model and estimation of its reliability measure. The prediction efficiency of any geospatial predictive model and its resultant output (i.e. proposed DSM) is usually estimated by using independent information posterior to map elaboration. Therefore, here where we have used a separate training set, the accuracy of the proposed DSM technique in Kermanshah Province was evaluated by calculating relative operating characteristics (ROC) which depict the capability of a binary algorithm system as its discrimination threshold changes (Fawcett, 2006b; Feizizadeh et al., 2014b; Sabokbar et al., 2014). We have also used a simple overlay technique to show the percentage of known changes of EVI in various sensitivity classes.

Considering the ROC method, the area under the ROC curve (AUC) values, varying between 0.5 and 1.0, is used to evaluate the accuracy of the DSM. The AUC defines the accuracy of the proposed probabilistic model through describing the model ability to reliably predict the occurrence or non-occurrence of an event. The ideal model shows an AUC value close to 1.0, whereas a value close to 0.5 indicates inaccuracy in the model (Fawcett, 2006b; Roodposhti et al., 2014b; Shahabi and Hashim, 2015b). In order to apply the ROC method, a concise and representative dataset was prepared using randomly selected 278 first negative along with 278 first positive points of EVI change throughout the study area. As it is assumed that changing EVI values are representative of drought impacts on vegetation cover, these points will be further used to evaluate the accuracy of sensitivity maps proposed

by two OC-SVM algorithms. The AUC value of ROC curve for the output map was found to be 0.809, with an estimated standard error of 0.01 (Fig. 5.11).

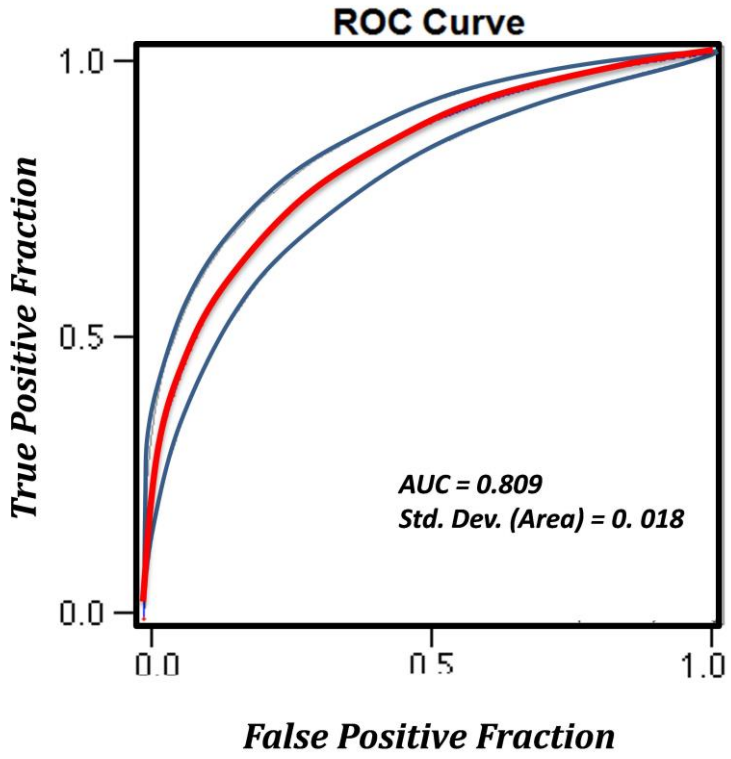


Fig. 5.11. ROC curve for the proposed DSM using two OC-SVM class.

The DSM results were also verified using the test inventory data points itself. Accordingly, these 278 first negative along with 278 first positive points of EVI change locations were overlaid on the proposed DSM map (Fig. 5.12). The result shows that 197 data points which represent vegetation lost (70.86 percent of all test data for negative changes) overlaid on the high and very high sensitive patches, which they only cover 14657 km² (59.1 percent) of the study area while only 1 test data appears in the very low sensitive zone.

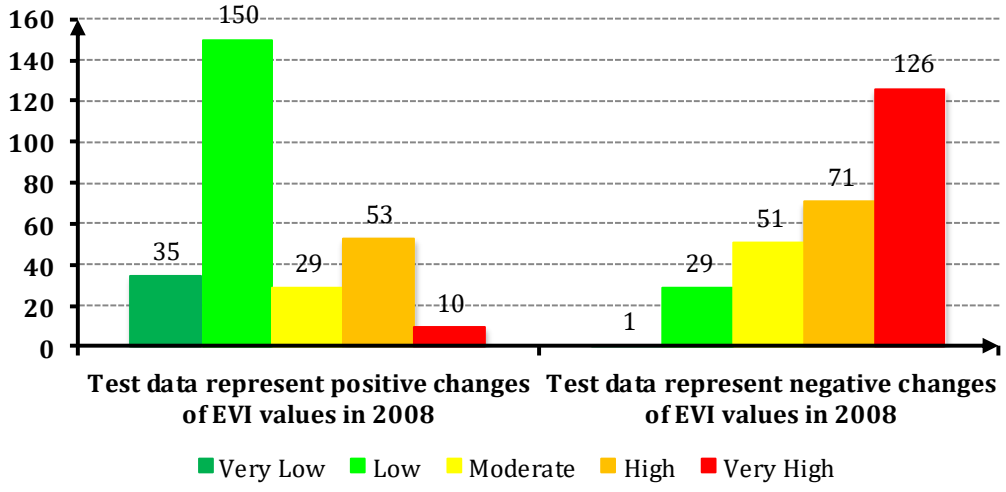


Fig. 5.12. Histogram of test data overlay showing the relative areas for each sensitivity class (each class is labeled with the number of the observed negative or positive test data accordingly).

In addition to the above, it should be mentioned that about 185 (66.5 percent of all test data for positive changes of EVI value) data points which represent positive change in EVI value overlaid on the low and very low sensitivity zones, which they only cover 3963 km² (15.9 percent) of the study region while only 10 test data appears in the very high sensitive class.

5.5. Discussion and conclusions

Considering the most important factors which cause the present conditions of sensitivity to drought being in place in the study region, there was a demand to conduct a reliable DSM for vegetation cover. The reliability of DSM depends not only on the presence of concise and representative database, in terms of data scale and accuracy, but also on the selection of the appropriate DIs for and drought identification and severity evaluation. Regardless of data scale and accuracy, the present study aimed to explore DSM of Kermanshah by using both climatic and satellite-based DI which were used to confirm mild and severe drought occurrence and resultant vegetation loss. Further, an objective classification scheme along with four different drought-related geo-data layers were used in order to produce a reliable DSM.

Results of the analysis showed notable spatio-temporal variation in terms of vegetation sensitivity to drought phenomena throughout the study area. In this respect, the elevation class of 1500 to 2000 meter, slope classes between 4 to 32 degree, aspect classes of south and south-west (May) or east and north-east (October) and Zagros Orogenic Belt are the most affected from drought periods. However, slope classes between 2 to 4 degree (both in May and October) or more that 32 degrees (October), elevation less than 1500 meter (both in May and October) or more than 2500 meter (October) and Sanandaj-Sirjan geomorphic zone are the least affected by drought periods.

According to the obtained results, the resistance of vegetation to drought occurrence may be mediated through the effects of topography (elevation, slope and aspect) on soils and microclimate. In this respect, the obtained results of proposed DSM testify that plants in higher elevations (i.e. above 1500m to 2500m) typically are more affected by drought, compared with the similar plants of lower elevation (below 1500m). This is mainly correspondent to the negative correlation between slope gradient and soil moisture which is statistically more significant in higher elevations where steep slopes prevail. On the other hands, depth of soil horizons is significantly increased by a decrease in elevation. Accordingly, soils at lower elevations of the slope (i.e. Cambiosol and Litosol) may have larger sources of water compared to the soils in higher elevation. Natural vegetation cover is relatively drought-tolerant at elevations above 2500m because of consistent snow coverage which increases soil moisture and decreases environmental temperature.

Consequently, the spatial variation in soil moisture is also controlled by slope angle and aspect. Here, drought impacts on vegetation loss more obvious within southern and eastern slope and the slope angle more than 8 degrees while it is less obvious in northern and western slope and the slope angle between the 0-4. The magnitude soil moisture was found to decrease not only with an increase in angle of slope and but also with a topographic solar radiation index derived from slope angle and slope aspect.

Considering geomorphic units criterion, it has been found that Sanandaj-Sirjan unit is less sensitive to the occurred drought events compared to different parts of Zagros Orogenic Belt (i.e. Zagros Thrust or Folded Zagros zone). This is mainly related to some primary variables such as the surface materials and elevation which further affects other secondary variables including temperature, evapotranspiration and snow accumulation coefficient.

As a matter of fact, this study not only presents an integrated strategic DSM framework with an emphasis on solving the decision problem by using an objective procedure. In other words, this article introduces the use of the enhanced vegetation index (EVI) as an indicator of soil moisture while it focuses on producing a drought sensitivity map (DSM) for vegetation cover using two one-class support vector machine algorithms. Finally, considering the fact that, the proposed schematic framework has the advantage of satellite imagery archive as ground truth; it can be used for production DSM of different local or regional scales. On the other hand, the proposed DSM scheme can be further

extended by using more drought-related criteria including vegetation type, soil depth, drainage density etc. through different case studies.

6. Fuzzy Shannon entropy: a hybrid GIS-based landslide susceptibility mapping

Chapter 6 develops an optimal solution to deal with spatial data uncertainty for both criteria standardisation and prioritisation that is applicable as a numerical solution for susceptibility mapping using multi-criteria models while there is no sufficient knowledge about the area of interest. This chapter is an attempt to achieve objective number five through modelling uncertainty Type B. The current chapter has been published in **Entropy** journal on 27th September 2016..

Shadman Roodposhti, M., Aryal, J., Shahabi, H. and Safarrad, T., 2016. Fuzzy shannon entropy: a hybrid GIS-based landslide susceptibility mapping method. Entropy, 18(10): 343.

Abstract: Assessing Landslide Susceptibility Mapping (LSM) contribute in reducing the risk of living with landslides. Handling the vagueness associated with LSM is a challenging task. Here we show the application of hybrid GIS-based LSM. The hybrid approach embraces fuzzy membership functions (FMFs) in combination with Shannon entropy, a well-known information theory-based method. Nine landslide-related criteria along with an inventory of landslides containing 108 recent and historic landslide points are used to prepare a susceptibility map. A random split into a train (≈ 70%) and test (≈ 30%) samples are used for training and validation of the LSM model. The study area – Izeh- is located in the Khuzestan province of Iran a highly susceptible Landslide zone. The performance of the hybrid method is evaluated using receiver operating characteristics (ROC) curves in combination with area under the curve (AUC). The performance of the proposed hybrid method with AUC of 0.934 is superior to multi-criteria evaluation approach using a subjective scheme in this research in comparison with previous study using the same dataset through extended fuzzy multi-criteria evaluation with AUC value of 0.894, and was built on the basis of decision makers' evaluation in the same study area.

6.1. Introduction

Landslide is either geophysical or climate-related disaster that is described as mass movements of earth surface material. This usually involves shear displacement of soil and/or rock masses along one or several slip surfaces (Varnes, 1978). Landslide susceptibility map (LSM) is a promising solution for both understanding and predicting probable future landslides. It assists planners in decision-making phase aimed for further mitigation of landslide consequences. Accordingly, a LSM depicts areas likely to have landslides in the future by correlating some of the principal factors that contribute to landslides with the past distribution of slope failures (Brabb, **1984**). In this respect, production of LSM at the early stage of landslide assessments is of crucial importance for safe economic planning, such as urbanization activities and the engineering of structures. However, a standard procedure for the production of landslide susceptibility maps does not exist (Ercanoglu and Gokceoglu, 2004).Thus, LSM can be accomplished by providing risk managers with easily accessible, continuous, and accurate information about landslide occurrence. The predictive capacity is poorly understood in LSM and is vague. In general, the spatial prediction of landslides is not easy due to the complex nature of landslides (Bui et al., 2012). LSM provide important information for predicting landslides hazards which include an indication of the time scale within which particular landslides are likely to occur (Atkinson and Massari, 2011). The associated vagueness can be dealt using Fuzzy sets theory.

Introduced by Zadeh (1965), Fuzzy set theory handles indefiniteness arising from intrinsic ambiguity than from a statistical variation (Zadeh, 1965b). A functional defined on the class of generalised characteristic functions (fuzzy sets), called "entropy", is introduced using no probabilistic concepts in order to obtain a global measure of the indefiniteness connected with the situations described by fuzzy sets (De Luca and Termini, 1972). The meaning of this quantity is quite different from the one of classical entropy because no probabilistic concept is needed in order to define it. This function gives a global measure of the "indefiniteness" of the situation of the problem at hand (Kosko, 1986). Although there is a well-defined mathematical theory of probability, there is no universal

agreement about the meaning of probability. Thus, for example, there is the view that probability is an objective property of a system and another view that it describes a subjective state of belief of a person. Then there is the frequentist view that the probability of an event is the relative frequency of its occurrence in a long or infinite sequence of trials. Thus, entropy is often used as a characterization of the information content of a data source, this information content is not absolute: it depends crucially on the probabilistic model (Majerník, 2014).

Effective LSMs could provide a proper understanding of "susceptible regions" (Xu et al., 2012). In order to better assist planners in understanding landslide hazard, a variety of GIS-based susceptibility mapping techniques are employed and developed (Shahabi et al., 2015). These approaches can be classified into three main groups such as subjective, objective and hybrid methods. The subjective methods typically include inventory mapping and decision makers' (DMs) evaluation in both standardisation and weighting of selected criteria (Wang et al., 2012). There are various GIS-based studies on LSM through the use of subjective approaches. Some of them used multi-criteria evaluation (MCE) techniques including: simple additive weighting (Feizizadeh and Blaschke, 2013), ordered weighted average (Ayalew et al., 2004), analytical hierarchy process (Yalcin, 2008), analytical network process (Neaupane and Piantanakulchai, 2006), PROMETHEE (Roodposhti et al., 2014c) etc. and some used different heuristic and knowledge driven techniques in order to assess landslide susceptibility mapping (Barredo et al., 2000; Ruff and Czurda, 2008; Van Westen and Getahun, 2003). Other studies, on the other hand, have shown a variety of objective methods in the assessment of the landslide susceptibility because of some limitations such as insufficient knowledge about the area of interest. The objective methods mostly rely on statistical (Budimir et al., 2015; Greco et al., 2007; Havenith et al., 2006; Lee and Min, 2001; Nandi and Shakoor, 2010; Ozdemir, 2011; Shahabi and Hashim, 2015a), soft computing (Bui et al., 2012; Lee et al., 2004; Pradhan and Lee, 2010), deterministic analysis (Carrara, 1983), neuro-fuzzy (Bui et al., 2012; Pradhan, 2013), artificial neural network (Bui et al., 2016; Conforti et al., 2014; Dou et al., 2015), decision trees (Hong et al., 2015; Yeon et al., 2010), and index of entropy (Constantin et al., 2011; Devkota et al., 2013a; Pourghasemi et al., 2012a; Tsangaratos and Ilia, 2016; Youssef et al., 2016), which are more rigorous and mostly relying on objective assessments. On the other hand, there are various hybrid GIS-based LSM methods which are both subjective and objective. In other words, some hybrid GIS-based LSM methods used subjective standardisation and an objective weighing technique (Davis and Blesius, 2015; Feizizadeh et al., 2014a; Zongji et al., 2010), and vice versa.

The accuracy of LSM mostly depends on the amount and quality of available data, the working scale and the selection of the appropriate methodology of analysis and modelling (Roodposhti et al., 2014c). In methodology implementation and its assessment, landslide casual criteria play a key role. In this study, we decipher the optimality of predictive solutions for objective criteria weighting. In an attempt to find an optimal solution, we show how modified Shannon entropy algorithm by fuzzy set theory can be successfully applied to the numerical solution of the LSM while there is no sufficient knowledge about the area of interest. In other words, the main objective of the present study is to extend a hybrid GIS-based LSM method within which fuzzy membership functions (FMFs) have been applied for criteria standardisation using "global knowledge" about landslides, while no "local knowledge" is utilised for criteria weighting. In literature, although different GIS-based models have been used for landslide susceptibility mapping, but LSM map extracted from modified Shannon entropy algorithm by fuzzy set theory has seldom been carried out. Therefore, this study aims to fill this gap in the relevant literature.

Since the LSM deals with a various sets of criteria it can be assumed that integration of fuzzy set theory with information theory, and in particular with Shannon entropy, will assist in performing accurate landslide susceptibility mapping. This accurate LSM is due to the flexibility of fuzzy membership functions and objective evaluation of criteria weights. Based on this assumption, the present research is an attempt to propose a novel hybrid method, which contributes to the objective decision making for regional landslide management. In other words, by using only the entropy values of previous landslide events for each criterion and regardless of experts' opinions, we intend to

facilitate criteria weighting process while improving or preserving LSM predictive accuracy compared with accurate subjective methods.

The paper is organised as follows. After a description of the study area in section 6.2, a detailed definition of the material and methods of the research is described in section 6.3. Section 6.4 presents results while section 6.5 discusses the achieved results and contributions, respectively. At the end, we provide the conclusions of this research in section 6.6.

6.2. Description of study region

Izeh is located in the eastern part of Khuzestan province, located in south-western Iran (see Figure 6.1), where, the high susceptibility for a mass movement and in a particular landslide is considered as a potential natural hazard for human society and their activities such as hydropower plants in Izeh. According to the inventory of landslides by the ministry of natural resources (MNR, 2010), there are 108 recorded landslide events in the region.

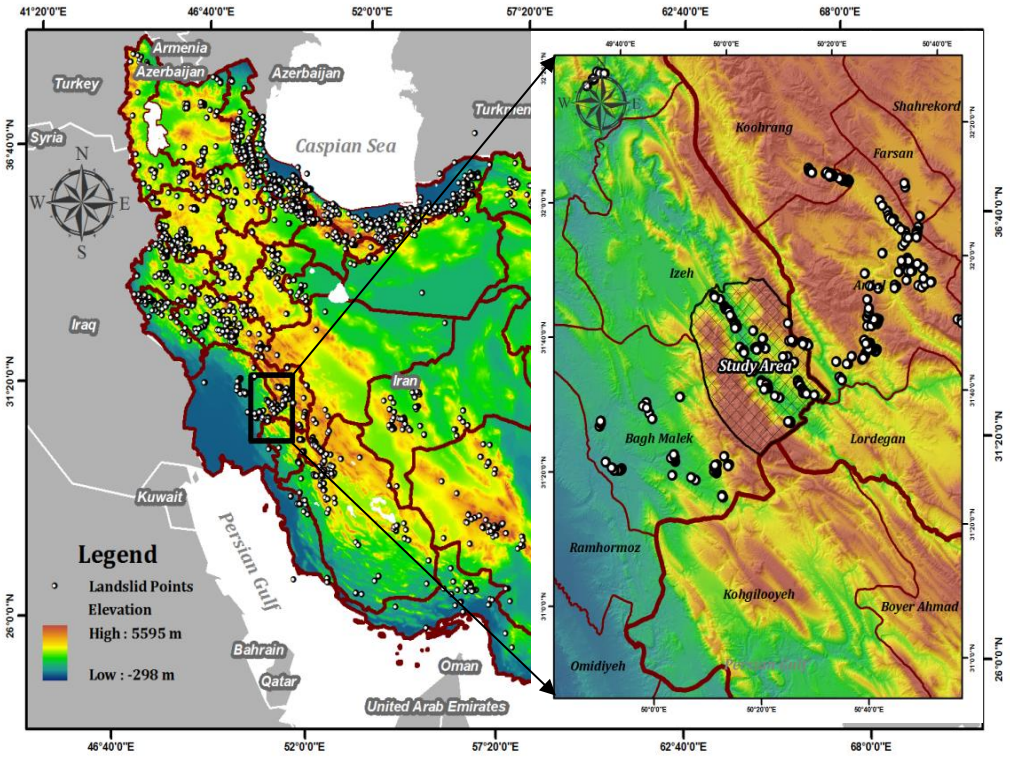


Figure 6.1. Location map of the study area

The climate is a temperate in north, while in south warm climate prevails. Similarly, mean annual precipitation within the study area varies from 450 to 700 mm. The region is important in terms of the agricultural activities and in particular hydropower plants. The Karun River, the main and longest river in all of Iran, passes through this area. The suitable topography of Karun canyon provides the possibility of constructing hydropower plants and three main dams have been constructed so far on different branches of the Karun River (Feizizadeh et al., 2014a).

Geologically, there are several minor faults and one major thrust in the region along with the 13 types of geologic formations cropping out in the region. The Izeh fault zone is a transverse fault zone with right-lateral strike slip (and some reverse component) in the Zagros Mountains, south-western Iran. That is majority controlled by the subsidence and sedimentation of the embayment. In terms of 13 types of geologic formations, nearly all of them composed of sedimentary rocks including, marl, shale,

limestone, gypsum, siltstone and other Quaternary deposits. It also should be mentioned that in the case of any triggering cause, there will be a significant chance of landslide occurrence within the south and south-east where the rough topography coincides with major thrust fault, Karun canyon and susceptible lithology. In other words, where there is susceptible lithology, proximity to faults contributes to slope instability, affecting not just surface structures but also terrain permeability. Eventually, the erosion associated with Karun River in nearby areas further leads to slope instability and generally increases the rate of subsequent slope failure. This is considered another prominent reason for the notable landslide recurrence in the region (Feizizadeh et al., 2014a).

6.3. Materials and Methods

6.3.1. Landslide influencing data layers

First of all, with respect to the available peer-reviewed GIS-based LSM research, nine criteria of the study area have been employed and prepared (Table 6.1 and Figure 6.2).

Table 6.1. Selected landslide related criteria on the basis of literature review

Criteria	Data source	Former studies using the same criterion for GIS-based LSM
Slope	30m, STRM DEM	(Lee and Min, 2001); (Komac, 2006); (Ayalew et al., 2004); (Conoscenti et al., 2008); (Thiery et al., 2007); (Yalcin, 2008); (Kayastha et al., 2012); (Bennett et al., 2013); (Kritikos et al., 2015)
Aspect	30m, STRM DEM	(Ayalew and Yamagishi, 2005); (Komac, 2006); (Guzzetti et al., 2006); (Thiery et al., 2007); (Yalcin, 2008); (Lotfi and Fallahnejad, 2010)
River	1:50,000Topo-map	(Yalcin, 2008); (Feizizadeh et al., 2014a); (Sabokbar et al., 2014)
Drainage	1:50,000Topo-map	(Yalcin, 2008); (Pareek et al., 2010); (Roodposhti et al., 2014c); (Feizizadeh et al., 2014a)
Fault	1:100,000,Geo-map	(Havenith et al., 2006); (Kanungo et al., 2006); (Lee and Pradhan, 2007); (Marjanović et al., 2011); (Shahabi et al., 2013)
Rainfall	30 years IMO data	(Hong et al., 2005); (Guzzetti et al., 2007); (Feizizadeh et al., 2014a)
Road	1:50,000Topo-map	(Ayalew and Yamagishi, 2005); (Yalcin, 2008); (Youssef et al., 2009); (Bathrellos et al., 2009); (Pradhan, 2013)
Lithology	1:100,000 Geo-map	(Ercanoglu and Gokceoglu, 2004); (Ayalew and Yamagishi, 2005); Thiery et al. 2007(Thiery et al., 2007); (Akgun et al., 2012a); (Davis and Blesius, 2015)
Land use	30 m, Landsat image	(Lee and Pradhan, 2007); (Bathrellos et al., 2009); (Feizizadeh et al., 2014a)

It must be noted that landslide susceptibility map of the study area has been derived from landslide related criteria mentioned in above Table 1. Road, river and drainage input maps were

extracted from the topographical map (1:50,000) of the study area, while the fault and lithology maps were obtained from geologic maps (1:100,000). In addition, the slope and aspect criteria were derived from 30m shuttle radar topography mission (SRTM) digital elevation model (DEM). Land use/cover maps were derived from Landsat ETM+ satellite images with 30m spatial resolution employing image analysis methods (MANR, 2008).

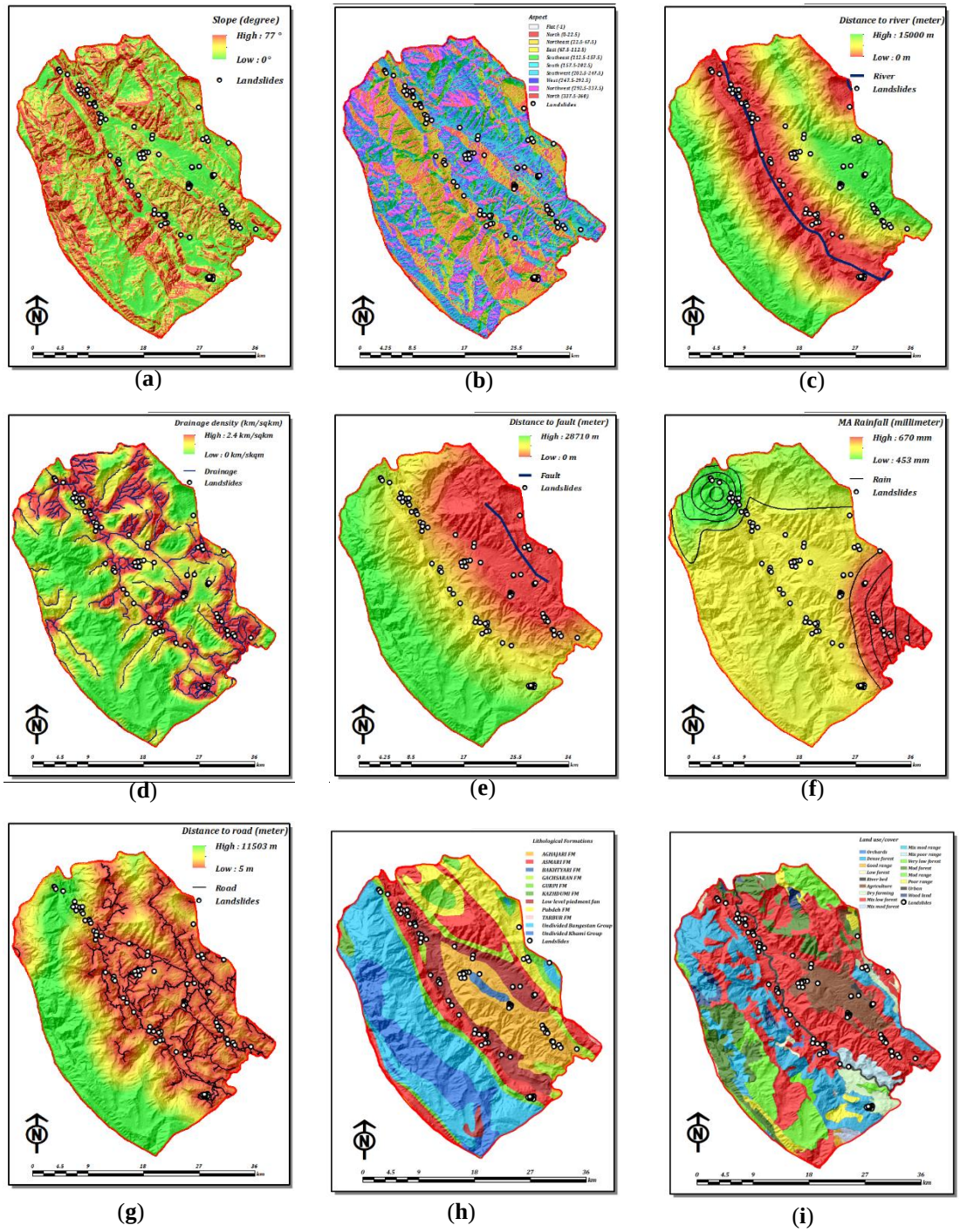


Figure 6.2. Nine applied criteria used in LSM of Izeh involving: (**a**) Slope; (**b**) Aspect; (**c**) Distance to river; (**d**) Drainage Density; (**e**) Distance to faults; (**f**) Mean annual rainfall; (**g**) Distance to roads; (**h**) Lithology and (**i**) Land use/cover

The average of 30 years mean annual rainfall data of Iran Meteorological Organization (IMO) was used to create mean annual rainfall map using kriging interpolation methods in ArcGIS environment. Finally, we also used an inventory of landslides containing 108 recent and historic landslide points which were recorded by GPS in field survey (MNR, 2010) for both geo-data layer weight evaluation and further validation of proposed LSM. The recorded landslides points are centroids of of each landslide polygon. Almost all of these landslides belong to slide-type landslides which are down-slope movement of material along a distinctive surface of weakness such as a fault, joint or bedding plane. In terms of landslide inventory, it was randomly split into a train (≈ 70%) and test (≈ 30%) samples for training the proposed hybrid model and subsequent validation purpose, prior and posterior map elaboration, respectively.

6.3.2. Proposed methodology

In order to depict the proposed methodology, it is best to consider a three-step procedure: in step 1, using fuzzy sets theory, data standardisation has been implemented in ArcGIS environments. To this end, a proper FMF is fitted on each selected criterion posterior to preprocessing phase. These FMFs has been selected according to peer literature review of similar LSM studies and local expert opinions. Accordingly, in step 2, the Shannon entropy is used for further evaluation of criteria weights, which determines the subsequent contribution of each landslide related criteria in overall susceptibility. This phase is implemented in MATLAB. Here, Shannon entropy technique is used as an objective-weighting scheme in LSM process. Finally, in the third step, results from above two steps are integrated using ArcGIS software (Figure 6.3). Further, analysis results were validated using receiver operating characteristics (ROC) curves and simple overlay technique using MATLAB and ArcGIS environments, respectively. Each step is explained as below:

6.3.2.1. Fuzzy membership function (FMF)

A major contribution of fuzzy set theory and related fuzzy membership functions (FMFs) is its capability of representing vague data. The theory also allows mathematical operators and programming to apply to the fuzzy domain. A fuzzy set is a class of objects with a continuum of grades of membership (Zadeh, 1965b). Such a set is mainly characterised by a membership function, which designates a membership value to every single object ranging from 0 and 1 and vice versa. In terms of LSM, fuzzy sets approve the possibility of partial membership of a considered geographic location to more than one susceptibility class. FMFs accordingly, were used to determine variation pattern forming gradual class boundaries between each susceptibility class. The shape of each applied FMF determines how the transition between 0 and 1 takes place.

6.3.2.2. Shannon entropy

The information entropy application originated from thermodynamics in 1948 (Zongji et al., 2010). In information theory, entropy is the quantitative measure of system disorder, instability, imbalance, and uncertainty and can forecast development trend of specified system (Lotfi and Fallahnejad, 2010; Sujatha, 2012; Yufeng and Fengxiang, 2009; Zongji et al., 2010). The Shannon entropy usually indicates to quantification of the expected amount of information enclosed by a message. At present, the information entropy method has been widely used to determine the weighted index in natural hazards, and in environmental, integrated assessment of natural processes such as debris flow, drought, sandstorm, etc. (Mon et al., 1994; Wang et al., 2011).

In terms of the landslide, it measures the dissimilarity or diversity in the environment, indicating the potential of each factor in causing landslides. In other words, the entropy of landslides refers to the extent the various factors influence landslide. Greater is the entropy index, greater is the influence of the factor in causing landslide (Sujatha, 2012). Finally, it also should be mentioned that various landslide related criteria are not the same regarding their attributes and dimension. Therefore, is not possible to conduct a direct comparison between those mentioned criteria which are applied in a LSM

process. In order to construct proper comparison, it's necessary to conduct standardisation process in the first step (See Equation 6.1 and 6.2).

$$\text{Positive effect} = \begin{cases} 1 & x = x_{\max} \\ 0.5(1 - COS(\pi \frac{x - x_{\min}}{x_{mac} - x_{\min}})) & x_{\min} \prec x \prec x_{\max} \\ 0 & x = x_{\min} \end{cases} \quad (6.1)$$

$$\text{Negative effect} = \begin{cases} 1 & x = x_{\min} \\ 0.5(1 + COS(\pi \frac{x_{mac} - x}{x_{mac} - x_{\min}})) & x_{\min} \prec x \prec x_{\max} \\ 0 & x = x_{\max} \end{cases} \quad (6.2)$$

Equation 6.1 is applicable for specific criteria with positive effect on probability of landslide occurrence (such as drainage density and mean annual rainfall). It means the more the value of the considered criteria is, the more the probability of landslide is as a simple rule. However, for some other criteria (viz. distance to river, distance to faults and distance to road) equation 6.2 is well-suited, where the reverse condition exists. Then, landslide entropy matrix R is formed by m landslide samples and n geo-data layer:

$$R = \begin{Bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & r_{m,n} \end{Bmatrix} \quad (6.3)$$

Accordingly, Shannon entropy is defined by (Zongji et al., 2010):

$$E_j = -k \sum_{i=1}^{m} p_{i,j} \ln p_{i,j} \quad (6.4)$$

where E_j is entropy value, $p_{i,j}$ is value of *ith* landslide in *jth* criteria and *k* is a positive constant, essentially a choice of unit of measure which is given by:

$$k = (\ln m)^{-1} \quad (6.5)$$

where *m* is number of occurred landslide events. Accordingly, normalized decision matrix $p_{i,j}$ for each landslide criteria can be defined by:

$$p_{i,j} = \frac{r_{i,j}}{\sum_{i=1}^{m} r_{i,j}} \quad (6.6)$$

The weights have attributed the role the factors play in the synthesis assessment, and the bigger value indicates that the factor's function is more important in this index system.

$$w_j = \frac{v_j}{\sum_{i=1}^{m} v_j} \quad (6.7)$$

where W_j is weight of *jth* geo-data layer and V_j is defined by:

$$v_j = 1 - E_j \tag{6.8}$$

6.3.2.3. Hybrid landslide susceptibility mapping model

The proposed hybrid model of landslide susceptibility mapping can be defined as:

$$S = \sum_{i=1}^{n} w_i \otimes x_i \tag{6.9}$$

where S is a degree of landslide susceptibility, W_i stands for the weight of each criterion and X_i is standardised landslide criteria.

6.3.3. Methodology implementation

The following 3-step experimental design is implemented (See Figure 6.3).

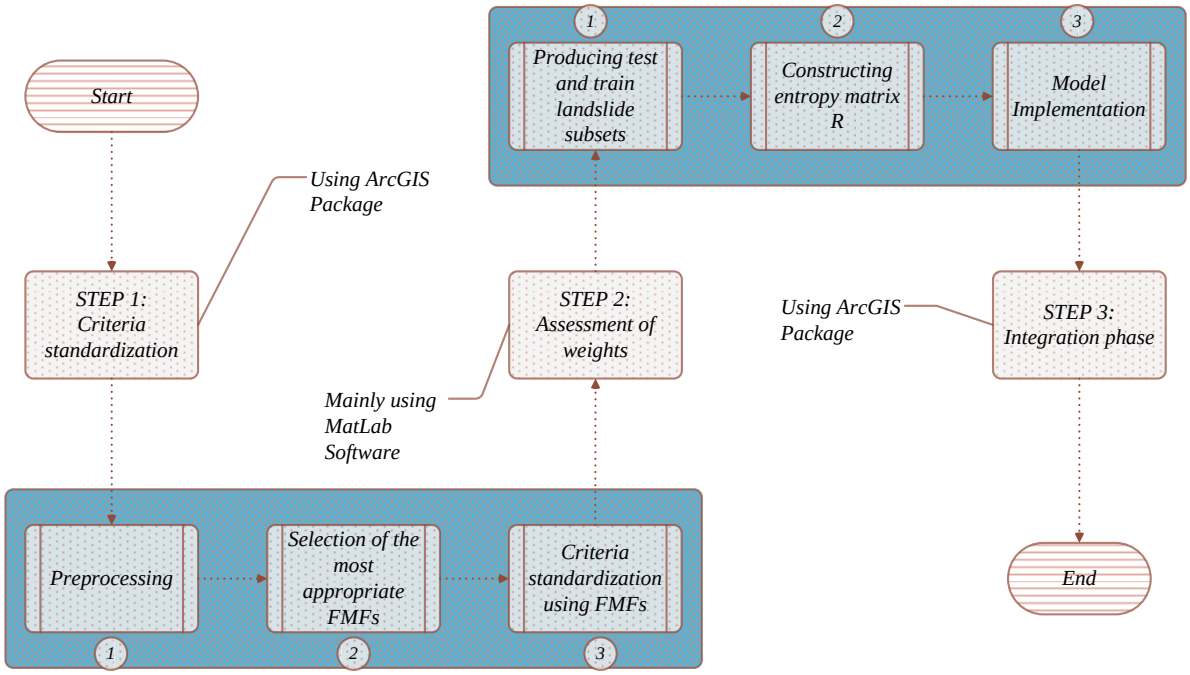


Figure 6.3. Schematic representation of the 3-step methodology implementation

6.3.3.1. Step 1: Data standardisation using FMFs

Considering the fact that GIS-based landslide related criteria measured not only in different units but also in different scales of measurement, such as nominal, ordinal, interval, and ratio scales (Akgun and Türk, 2010b), there is an urgent need for data standardisation. This rises from the inherent need to integrate all landslide criteria into the single output in the evaluation process. In this regard, the fuzzy membership approach is considered one of the frequently applied standardisation methods that have been proposed (Llu et al., 2004).

The use of fuzzy sets within GIS-based hazard and susceptibility assessment has been demonstrated to have a good effect (Akgun and Türk, 2010b; Llu et al., 2004; Mason and Rosenbaum, 2002). For this reason, fuzzy sets were used in this study. In this context, all the factors used were standardised to a float-level range of 0–1, where 0 is assigned to the least susceptible areas and 1 to the most susceptible ones. This transforms the different measurement units of all landslide casual criteria

into comparable values using FMFs (Gemitzi et al., 2006). Figure 6.4 shows selected and further applied FMFs for LSM of the study region.

There is no optimal method for choosing the most appropriate FMF and their respective parameters; these are generally selected according to the preferences of the DMs (Rojas-Mora et al., 2013; Roodposhti et al., 2014c). However, the predictive and causal value of landslide casual criteria seems more or less similar in most of the studies. In this study, 3 different membership functions have been employed for landslide susceptibility purpose including sigmoidal (s-shaped) FMFs, i.e., monotonically increasing and monotonically decreasing, user-defined fuzzy membership functions along with crisp membership functions are specified for each criterion (See Figure 6.4). The sigmoidal membership function is likely the most commonly used FMF in fuzzy set theory, and provides a gradual variation from non-membership (zero) to complete membership (one) (Akgun and Türk, 2010b; Eastman, 2003; Llu et al., 2004; Zadeh, 1965b), whereas it is sometimes inevitable to use user-defined FMFs or crisp membership functions. Nevertheless, all applied functions of criteria and the resultant output raster files are shown in Figure 6.4 and Figure 6.5 respectively.

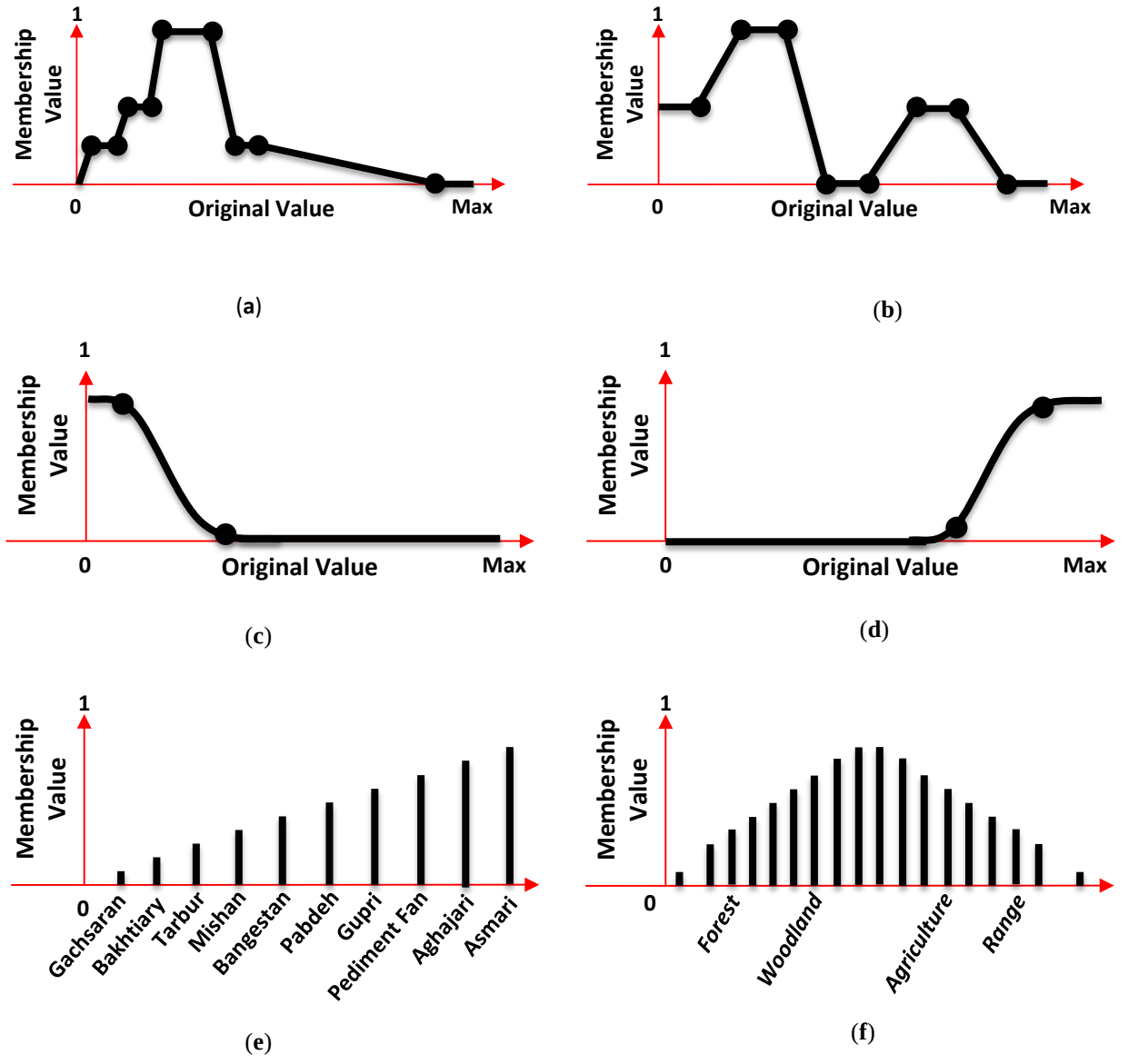


Figure 6.4. All types of used membership functions including: (**Type I**) User defined FMF (for: (**a**) slope and (**b**) aspect), (**Type II**) Sigmoidal FMF including both monotonically decreasing (for: (**c**) distance to river, distance to

faults and distance to road) and monotonically increasing (for: (**d**) drainage density and mean annual rainfall) and (**Type III**) Crisp MF (for: (**e**) lithology and (**f**) land use).

6.3.3.2. Step 2: Assessment of weights with Shannon entropy

Here, this study selects a total number of 76 landslides to calculate the weights of each landslide criteria in southern Izeh. Following the insertion of all 9 selected landslide criteria into the entropy matrix *R*, in order to standardise the basic data into the mentioned entropy matrix, both equations 6.1 and 6.2 were employed along with user defined FMF and crisp MF (Figure 6.4). Next, weights of all criteria are determined by the help of Shannon entropy in the successive steps.

$$R = \begin{matrix} 1 \\ 2 \\ \vdots \\ 76 \end{matrix} \begin{Bmatrix} 1.000 & 0.583 & 0.747 & 1.000 & 0.000 & 0.143 & 0.970 & 1.000 & 0.300 \\ 0.337 & 0.000 & 0.708 & 1.000 & 0.000 & 0.138 & 1.000 & 1.000 & 0.300 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.364 & 0.124 & 0.000 & 0.140 & 0.184 & 0.122 & 1.000 & 0.900 & 0.600 \end{Bmatrix}$$

Entropy values and weights can be calculated using (Equations 6.4-6.7):

$$E_j = \{0.883 \quad 0.874 \quad 0.735 \quad 0.902 \quad 0.715 \quad 0.861 \quad 0.915 \quad 0.927 \quad 0.902\}$$

$$v_j = \{0.116 \quad 0.097 \quad 0.264 \quad 0.097 \quad 0.284 \quad 0.138 \quad 0.084 \quad 0.072 \quad 0.097\}$$

$$w_j = \{0.090 \quad 0.097 \quad 0.206 \quad 0.075 \quad 0.222 \quad 0.108 \quad 0.066 \quad 0.056 \quad 0.075\}$$

Finally, after further calculation of entropy value, we obtained the weights (Table 6.2) to be used in criteria integration in the next step.

Table 6.2. The calculated weight vector from Shannon entropy method

Criteria	**Weight**
Slope	0.090
Aspect	0.097
Distance to river	0.206
Drainage density	0.075
Distance to Fault	0.222
Rainfall	0.108
Distance to roads	0.066
Lithology	0.056
Land use/cover	0.075

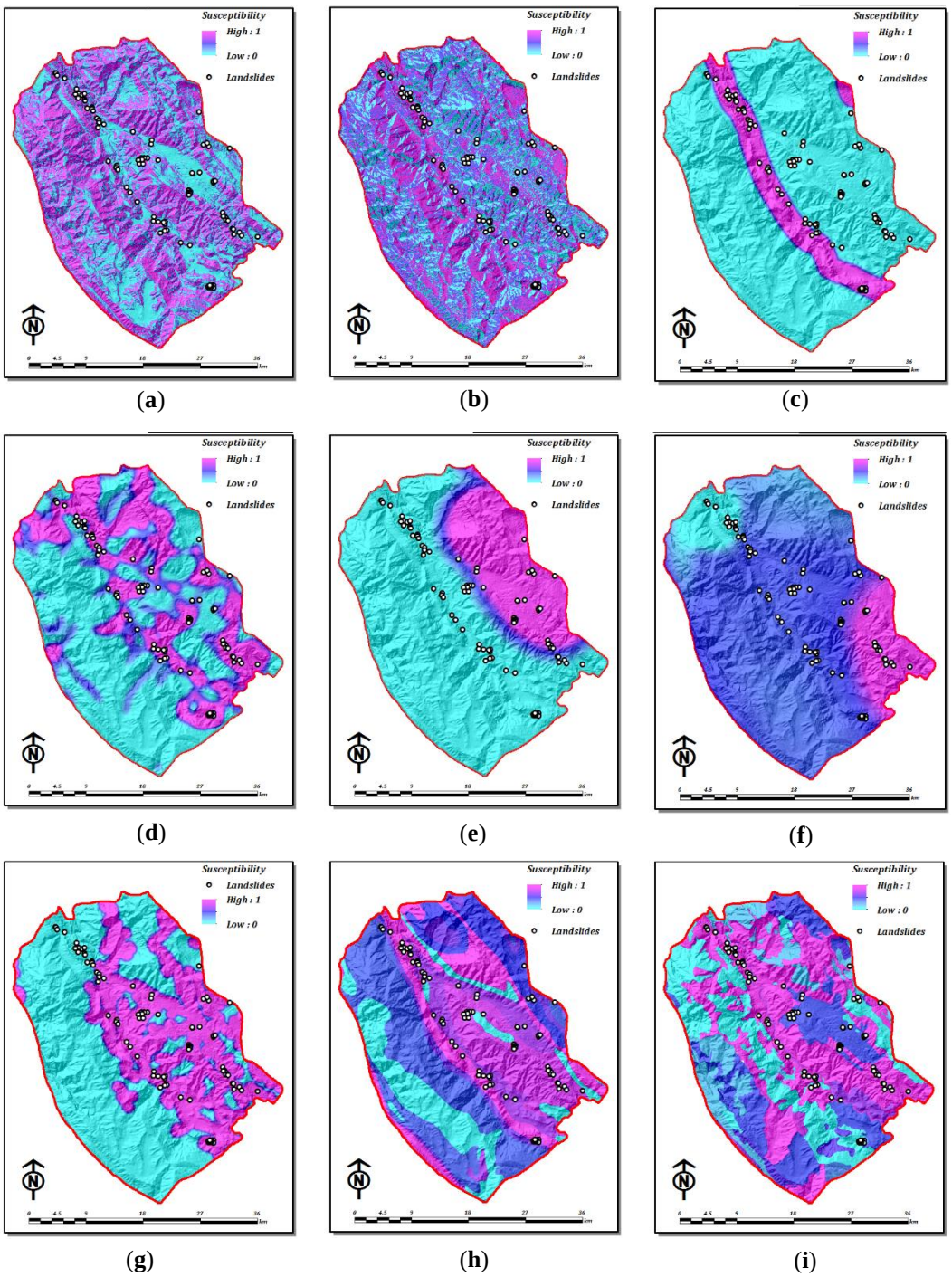


Figure 6.5. Obtained output after applying selected membership functions (i.e. Fuzzy or crisp) on each related parameter: (**a**) Slope; (**b**) Aspect; (**c**) Distance to river; (**d**) Drainage Density; (**e**) Distance to faults; (**f**) Mean annual rainfall; (**g**) Distance to roads; (**h**) Lithology and (**i**) Land use

6.3.3.3. Step 3: Integration phase

The prepared database of LSM is successfully georeferenced using Universal Transverse Mercator (UTM) coordinate system in ArcGIS environment. The weight derived from Shannon entropy index for each landslide related criteria is calculated using the 76 occurred landslide events and is applied for

integration purpose. Afterwards, the resultant susceptibility map is calculated as the summation of the weighted criteria as shown in Equation 6.9.

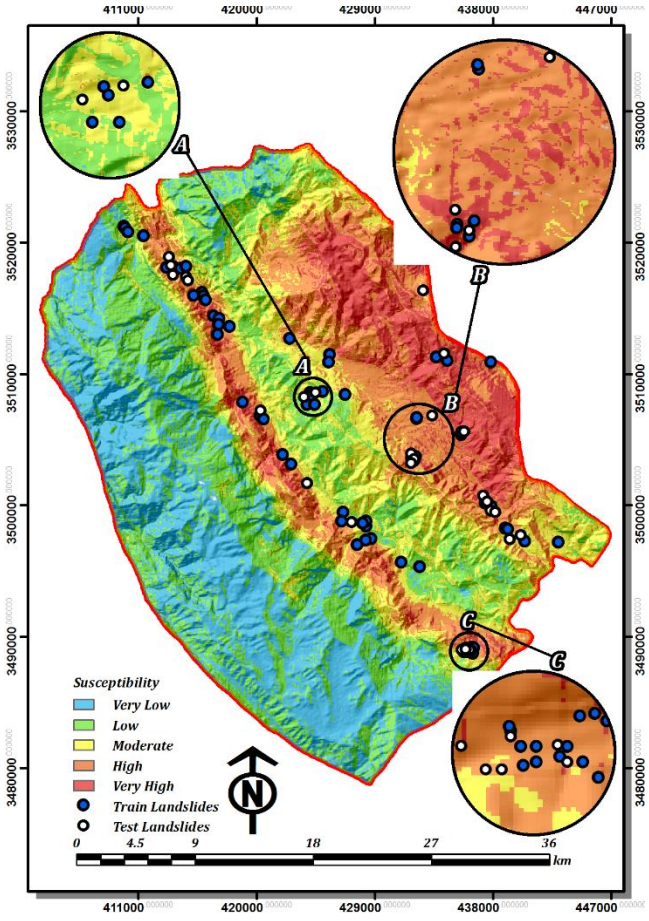


Figure 6.6. Final susceptibility map using the proposed hybrid GIS-based method. A, B and C circles are only for better representation of contiguous landslide points density and positions within the study area.

Using all the factors (Table 6.1), susceptibility values range from 0 to 8.00 showing various levels of susceptibility. The higher susceptibility values refer greater probability of expected landslides occurrence in the near future. Finally, the landslide susceptibility map (Figure 6.6) is divided into five susceptibility classes very low, low, moderate, high and very high using natural breaks classification.

The "natural breaks" classifier is based on natural grouping of data values. Normally, the breakpoints are identified by looking for groups and patterns inherent in the data. However, the reasons for using certain methods in previous works are usually not explained by the authors. In this study, the manual classifier method was used to reclassify the LSM values into five different susceptibility zones, according to the classification method that was proposed by (Galang, 2004).

6.4. Results

After layer standardisation of landslide casual criteria, the susceptibility map was produced based on a hybrid GIS-based LSM technique (See Figure 6.6). In terms of criteria weighting, typically, in spatial MCDA (multi-criteria decision analysis) problems the greater the value of the entropy corresponding to a spatial attribute, which implies the smaller attribute's weight, the less the discriminant power of that attribute in decision-making process (See Equation 6.8) (Lotfi and

Fallahnejad, 2010). Accordingly, fault and river criteria are considered as the two first important landslide criteria. Also, the distance from faults and river criteria both indicate potential trigger factors of the slope failure, are also among the principal indexes of a landslide. As a result, the objectively obtained weights of the landslide related criteria using Shannon entropy index is consistent with the basic rules of identification, characterization and development of landslides.

6.4.1. Validation of the results using ROC curve

The validation phase could be considered as one of the most fundamental stages in the development of all susceptibility maps and determination of their prediction capability for future usage in any natural hazards study and managements. The prediction efficacy of each LSM and its resultant output is typically evaluated by using available independent information of recorded landslide events, which are not used through LSM process (i.e. test subset of landslide inventory map) (Roodposhti et al., 2014c). As a result, in the present study, the landslide inventory database has been divided into two parts, including training and test datasets. Therefore, the accuracy of the proposed LSM in the study area was evaluated by calculating relative operating characteristics (ROC) (Fawcett, 2006a; Nandi and Shakoor, 2010; Roodposhti et al., 2014c) and percentage of known occurred landslides events in various susceptibility classes using test landslide samples. Here, the Area under the ROC Curve (AUC) value, ranging from 0.5 to 1.0, is a numeric indicator of map accuracy. Meaning that AUC is close to 1, the result of the test is more reliable, while closer the AUC to 0.5 indicates to the less reliable result (Pradhan, 2013; Roodposhti et al., 2014c).

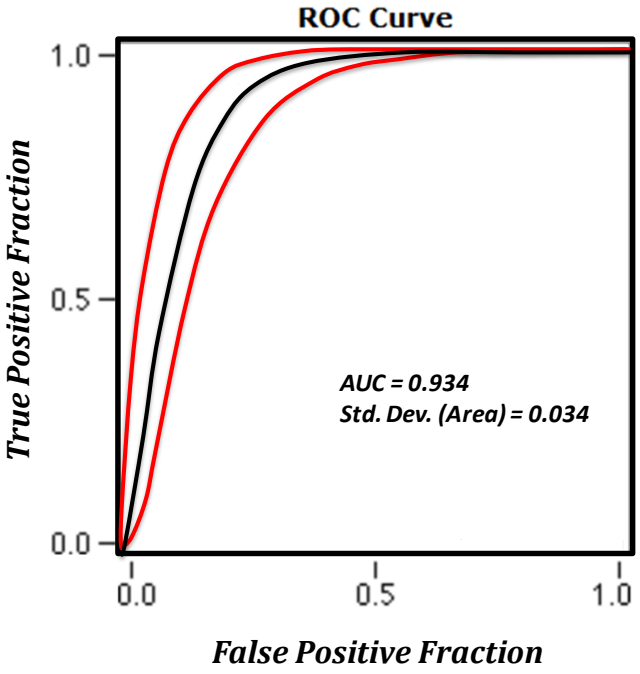


Figure 6.7. ROC curve for the proposed Landslide Susceptibility Map (LSM) of fuzzy Shannon entropy

In pursuance of further implementation of the ROC evaluation technique, a precise and comprehensive test dataset was prepared using 32 landslides and 32 randomly selected non-landslide points of the study area. In this regard, following the early identification of landslide-free area using aerial photo interpretation and field survey, non-landslide points are selected within the boundary of these landslide-free areas. Subsequently, the AUC value of 0.934 has been obtained with an estimated standard error of 0.022 (See Figure 6.7).

4.2. Validation of the results using simple overlay

In the second validation process, the LSM result has been evaluated using the test landslide locations, accordingly, these 32 points were overlaid on the susceptibility map of proposed hybrid GIS-based LSM (See Figure 6.6). The result shows that approximately about 90 percent of the recorded landslides occurred in the high and very high susceptibility classes, which they only cover 30.63 percent of the study area, while, there is no recorded landslide appears in the low and very low susceptibility zones. In addition to the above, only 3 landslide points (≈10% of all recorded landslides) fall into the medium susceptibility zone of the map which covers about 18.22 percent of the study area (See Figure 6.8).

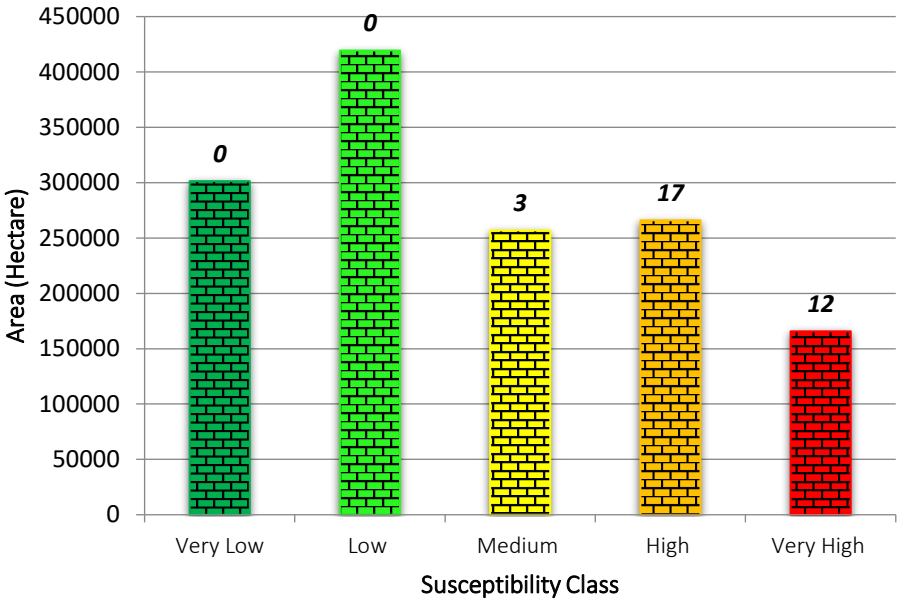


Figure 6.8. Histogram of calculated landslide susceptibility map showing the relative areas for each susceptibility class (susceptibility classes are labelled with the numbers of the observed landslide points accordingly).

6.5. Discussion

The accuracy of predictive models is considered as a major concern in the majority of environmental modelling applications including LSM (Bennett et al., 2013). The predictive accuracy of subjective LSM models can be affected by the inherent bias that emaciates from DMs' point of view during both data standardisation and criteria weighting. Moreover, the absence of expert DMs may make serious hindrance in LSM process while using a subjective method. Considering criteria standardization scheme, by applying a more computationally intensive approach we attempted to preserve the original quality of spatial data. In this respect using a variety of FMFs positively affected validity and accuracy of input spatial criteria. Therefore, missing or generalised values can represent in otherwise precise data. Further, the proposed methodology shows a promising result to predict landslide susceptibility values regardless of experts' opinion. According to the obtained results, the accuracy of proposed hybrid model is improved significantly compared with the accuracy of accurate subjective approaches, which have been previously implemented in the study area using the same dataset (Feizizadeh et al., 2014a).

6.5.1 Obtained results and relevance to the previous studies

Considering high frequency of landslide being in place in several areas of southern Izeh, there was a demand to conduct an accurate landslide susceptibility map. The expected accuracy of LSM depends not only on the presence of concise and comprehensive data, in terms of data scale and accuracy, but also on the selection of the appropriate methodology of data processing and modelling (Yalcin, 2008). Regardless of data scale and accuracy, the present study aimed to explore landslide susceptibility of southern Izeh by developing a hybrid GIS-based LSM that uses neither DM's evaluation nor sophisticated objective methods. This is an integrated strategic LSM framework with an emphasis on structuring the decision-making process problem. Within this approach, Shannon entropy was employed to determine the criteria weightings from an objective evaluation of spatial domain while different fuzzy membership functions were employed for criteria standardization.

Obtained results of ROC curve analysis (AUC=0.934) (See Figure 6.7) and simple overlay technique (See Figure 6.8) signify that the proposed hybrid fuzzy Shannon entropy evaluation technique is a promising tool for integrating multiple raster-based criteria for LSM while there is not sufficient knowledge about the criteria weights with respect to landslide mechanism of the study region. The previous study using the same dataset through extended fuzzy multi-criteria evaluation which was built on the basis of DMs' evaluation achieved AUC value of 0.894 (Feizizadeh et al., 2014a). This further approves the capability of proposed hybrid model for prediction of landslide susceptibility values. In other words, achieved results of accuracy metrics comparison approves that the proposed LSM model can achieve superior prediction accuracy to what that can be achieved by using DMs' points of view (Table 6.3), with significant time saving.

Table 6.3. Accuracy metrics of implemented data-driven (objective) and experts driven (subjective) LSMs using fuzzy Shannon entropy and extended fuzzy multi-criteria evaluation methods, respectively.

	Objective weighting approach	**Subjective weighting approach**
Number of Cases	64	212
Number Correct	56 (76% of total)	173 (81% of total)
AUC	0.93	0.89
Std. Dev. (Area)	0.01	0.02
Accuracy	76.6%	81.6%
Sensitivity	100.0%	98.1%
Specificity	53.1%	65.1%
Pos Cases Missed	0	2
Neg Cases Missed	15	37

6.5.2 Spatial information extraction and prediction

This study contributed in the area of a spatially structured dilemma of predicting landslide susceptibility values for specific geographic location. This may be implemented through standardising and subsequent summing of landslide casual criteria. In this paper, we attempted to present an assessment of LSM, carried out by the implementation of hybrid fuzzy Shannon entropy evaluation within which fuzzy set theory has been used for criteria standardisation, and Shannon entropy

algorithm was used for weighting of some factors that may affect the landslide susceptibility. Therefore, the prepared hybrid susceptibility map is the result of a pixel-based combination of nine standardised criteria affecting the degree of landslide susceptibility. The optimal criteria weights is obtained objectively by precise a mathematical solution through the proposed entropy-based model (Stein et al., 2005b). In this respect, the lower the landslide entropy of a criterion, the higher the weight is. In other words, the lower landslide entropy within certain criteria (i.e. distance to faults and distance to river) indicates the presence of predictive spatial frequency and vice versa.

Further, as expected, the estimated data driven (objective) criteria weights using Shannon entropy algorithm do not conform to the subjective criteria weights estimated using an aggregation of DMs' votes in our prior research (Figure 6.9).

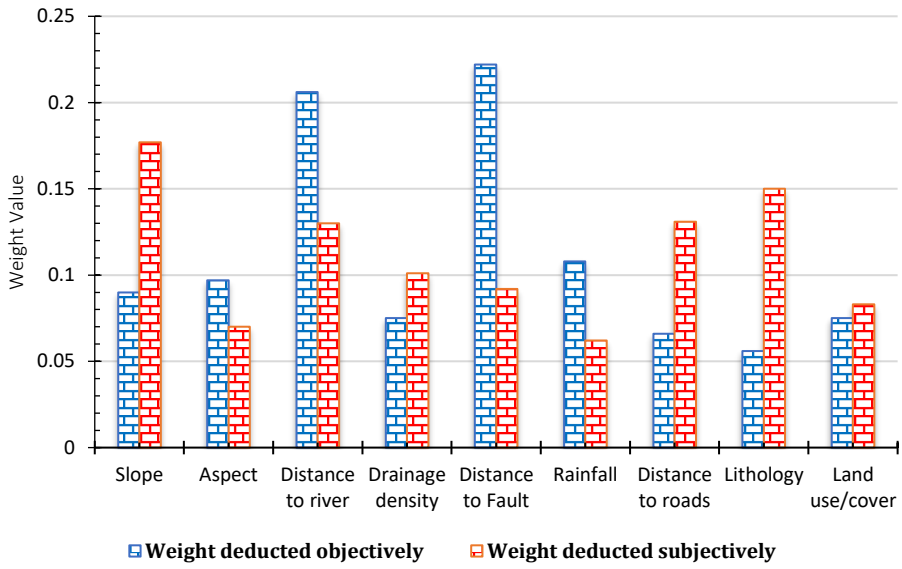


Figure 6.9. Histogram of estimated data driven (objective) and experts driven (subjective) landslide casual criteria weights for selected landslide criteria.

According to the obtained results of fuzzy Shannon entropy criteria weighting scheme distance to fault is the most important criteria followed by distance to river and rainfall criteria, respectively. Therefore, considering the estimated criteria weights, the spatial distribution of landslide susceptibility values is majorly controlled by these mentioned criteria. This may be further approved by high concentration of recorded landslide events alongside Karun River (Figure 6.6). Nonetheless, considering the DMs' evaluation slope is referred to as the most significant criterion followed by lithology and distance to road layers. Considering these two weighting approach, fuzzy Shannon entropy seems more realistic for predictive modelling of spatial pattern of landslides compared to the latter method. Even though the slope criterion is of paramount importance in any shape of slope instability, it is not the only constituent of landslides. Accordingly, the spatial pattern of landslides (at least in the study region) is controlled by important but less geographically available landslide casual criteria (distance to fault, distance to river and rainfall criteria). In other words, if similar high susceptible values of slope (or any other criteria) are prevailing all around a region while landslide distribution pattern is represented with different spatial order (Figure 6.5a), a secondary criterion (such as distance to river) with less availability may be determinant factor of landslide's spatial distribution (Figures 6.5c, 6.5e and 6.5f). This communicates the insight of the proposed objective weighting scheme in local evaluation of the landslide casual criteria. In other words, in the current study area, the slope angle is usually sufficient to influence landsliding. Nonetheless, considering the spatial distribution of landslides, there is limited evidence which approves the slope criterion plays an important role in landsliding. In the present study area, susceptible slope values are distributed almost evenly over the

study region; however, occurred landslide events are more or less concentrated along Karun Canyon. This may be due to the fact that the required water for slope failure, as a triggering factor, is controlling landslide events in southern Izeh. Water is not always directly involved as the transporting medium in mass movement processes while it does play an important rule. This is not only approved by obtained criteria weights of Shannon entropy method but also it can be recognized by visual inspection of landslides spatial pattern and frequency alongside Karun River (Figure 6.6).

Further, considering the results of our proposed objective weighting approach lithology criterion is the least important among all selected criteria while the expert opinion refers to the rainfall layer as the least important. The achieved accuracy value of fuzzy Shannon entropy, however, is yet remarkably superior.

6.5.3 Decision aiding and planning

Many researchers, (Castellan, 2013; Feizizadeh and Blaschke, 2013; Lerner et al., 2015) pointed out that the traditional subjective weighting schemes usually suffer from sensitivity in decision-making and they are susceptible to intrinsic errors of experts' knowledge. Looking into the contribution to decision aiding, this study presents an integrated strategic weighting procedure using an objective method which determines the criteria weights by solving mathematical models. This is executed without any consideration of the decision maker's preferences as it is a convention in subjective methods, such as the AHP method, OWA method, Delphi method and etc. In other words, this article introduces an objective approach that integrates fuzzy set theory and information theory algorithm (i.e. Shannon entropy), which could be a useful geospatial tool for integrating multiple features/attributes that affect the LSM process. This can largely compensate for the absence of expert DMs or the lack of local knowledge about study area when it comes to producing quality LSM.

6.5.4 Limitation of proposed methodology in LSM

While information theory-based methods such as the one proposed in the present research have shown a considerable potential in different predictive spatial modelling scenarios, they do have their own constraints. Even though the application of the proposed methodology as an objective weighting scheme is not dependent on decision maker's expertise and judgment, it relies on quantification of defined attributes of landslide data points using step by step mathematical computation. This is conditional on the existence of the concise and representative database. In terms of the present research, comprehensive and readily accessible landslide inventory database was quite beneficial in achieving the desired outcome. However, the inefficiency of a continued reliance on our small test sample size was particularly striking.

Another limitation of the implemented methodology is some false alarms in low slope areas as a high susceptible class (very few pixels as a very high susceptible class) within NE of the study region. This is mainly due to the fact that the slope angle is not characterised as a primary criterion shaping landslide occurrence spatial pattern. Compared to total landslide areas, most of LSM approaches end up with extremely high false positive rates in terms of high or very high susceptible areas. As this problem is not only limited to our study, therefore, exploring the ways to reduce mentioned problem may be a fertile ground to be addressed in future studies.

Further, posterior to fitting desired membership function, proposed fuzzy Shannon entropy technique considers the dataset as a collection of distributions, which may not be suitable to extract specific spatial structures embedded in the underlying features/attributes (Wang and Shen, 2011). Even though datasets with the same histogram certainly have the same entropy (i.e. distance to river and distance to fault in the present study), the distributions of their data values in space could be totally different. In addition, the result can be sensitive to the level of discretization caused by different membership functions (i.e. crisp or fuzzy) when using the histogram. We believe that more interests from researchers with larger sample size of data are vital for developing more robust entropy based LSM frameworks that can incorporate generalizable results.

6.6. Conclusions

This Fuzzy Shannon entropy integration showed a promising results for GIS-MCDA as it tackles two major limitations. Firstly, the inherent subjectivity which is emaciated from DMs' preference is diminished during criteria standardization phase. Secondly, intrinsic bias and probable errors of DMs' preferences corresponding to the subjective weighting approaches is also eliminated using proposed LSM model. This LSM approach involves a thoughtful selection and elaborative standardization of landslide casual criteria while weighting procedures are accomplished using an objective method. This is performed by constructing a mathematical approach without any consideration of the DMs' preferences from the beginning to the end of model implementation. Our results show that the integration of fuzzy sets with Shannon entropy can contribute to the production of landslide susceptibility maps with a reasonably high level of reliability. Finally, considering the fact that, the proposed hybrid method has the advantage of objective weight evaluation; it can be used not only in similar areas of geo-hazards risk assessment and mapping, such as land subsidence, earthquake and flood risk mapping but also in multi-hazard risk assessment for a further combination of risk elements. However, in order to apply the proposed objective weighting approach more generally by conducting different case studies, new hybrid models of GIS-based landslide susceptibility mapping shall be developed.

7. Towards automatic calibration of neighbourhood influence in cellular automata land-use models

The current chapter is aimed to develop a randomized rule detection (RRD) procedure to automate the calibration of the neighbourhood dynamics (N), which controls the key model behaviour of land-use change models (LUC). This chapter is an attempt to achieve objective number six through modelling uncertainty Type C. The current chapter has been submitted and is published in **Computers, Environment and Urban Systems** journal.

Shadman Roodposhti, M., Hewitt, R.J. and Bryan, B.A., 2020. Towards automatic calibration of neighbourhood influence in cellular automata land-use models. Computers, Environment and Urban Systems, 79: 101416

Abstract: Cellular Automata (CA) land-use models are widely used for understanding processes of land-use change (LUC). However, calibration of these models is a knowledge-intensive and time-consuming process. Although calibration of common driving factors such as accessibility (A), or suitability (S) is a relatively straightforward task, calibrating the neighbourhood dynamics (N), which controls the key model behaviour, is often very challenging. Here, building on the SIMLANDER modelling framework, we develop an automatic rule detection (ARD) procedure to automate the calibration of N. To demonstrate the performance of the tool, we simulated 15 years of urban growth in Ahvaz, Iran (2000-2015) using a wide range of different rule-sets. We evaluated calibration goodness-of-fit for each rule-set against a reference map by means of cross-comparison of multiple metrics using a ranking procedure. The ARD approach facilitates model calibration by allowing rapid identification of the optimum ruleset from a wide range of possible parameter settings, while the ranking procedure facilitates comparison of simulations using multiple metrics. The approach we present also helps to improve simulation accuracy with respect to manual calibration methods, and increases understanding of neighbourhood dynamics for the urban area studied. To encourage repeatability and transparency, our approach uses only open data and Free-and-Open Source Software (RStudio environment) and we provide our ARD scripts as an Appendix to this paper.

7.1. Introduction

Land-use change (LUC) is an issue of global concern and a major focus of sustainability research (Gao and Bryan, 2017; Veldkamp and Lambin, 2001; Yang et al., 2014). Models which capture the complexities of land-use systems are important for analysing the driving forces of LUC, and simulating future scenarios (Cheng and Masser, 2004; Kamusoko and Gamba, 2015). Until a few decades ago, spatially explicit LUC models were constrained by computational capacity (Lopez, 2014). However, recent advances in computer technology, analytical methods, and spatial data have led to important advances in spatial modelling, including *predictive* LUC models (Lopez, 2014; Santé et al., 2010). These models offer useful information about future change by projecting historical or hypothesised LUC tendencies to future dates.

One commonly-used LUC simulation approach is cellular automata (CA) (Batty and Xie, 1994; Ku, 2016; Santé et al., 2010; Tobler, 1979; Torrens, 2006). CA is a simple but effective method for predictive LUC modelling (Guan et al., 2011; Yang et al., 2008) utilising a bottom-up and discrete modelling paradigm, with dynamics generated by a set of simple local rules (i.e. transition rules) (Liu et al., 2014). Originally conceptualised in the late 1940s by Ulam and von Neumann (Santé et al., 2010; Von Neumann and Burks, 1966), CA are patterns of cells on a specified tessellation, which evolve through a number of discrete time steps according to a set of transition rules based on the state of individual cells and their neighbourhood. These transition rules are applied iteratively within the pre-defined neighbourhood and generate system dynamics over multiple time steps (Moreno et al., 2010; White and Engelen, 1997; Wolfram, 2002). In a LUC model, each cell can have one of i possible discrete states/values at a time, where i is land-use type, e.g. urban, forest, agriculture, water. In SIMLANDER, the model on which this work is based (Hewitt et al., 2013), a single land-use is modelled at a time, so i

takes value 1 (urban) or 0 (not urban). While much more complex multiple land-use configurations are frequently incorporated into these kinds of models (see e.g. Carter, 2018; Hewitt et al., 2014) with the aim of understanding the dynamic interaction of competing uses, simple single-land-use models like SLEUTH and SIMLANDER are unrivalled for rapid exploration and simulation of urban growth, pattern and form, as their enduring popularity attests (e.g. Clarke et al., 1997; Feng et al., 2018a; Feng and Tong, 2019; Xian and Crane, 2005). Specifically, these kinds of models are especially relevant for understanding the phenomenon of urban sprawl (e.g. Lagarias, 2012; Leao et al., 2004).

In SIMLANDER, as in most CA models, the state of a cell at time t+1 (future cell state) is defined by the state of the same cell and immediate neighbouring cells (N) at time t (present cell state) (Jokar Arsanjani, 2012; Moreno et al., 2010), as well as other factors known to influence land-use change such as proximity to infrastructure networks, known as accessibility (A), the underlying land suitability for development (S) (Barredo et al., 2003; Barreira González et al., 2015; White and Engelen, 1993), and a random factor (v) to account for unknown or undetermined factors that influence real-world outcomes (Hewitt and Díaz-Pacheco, 2017; Hewitt et al., 2014). Although, all these factors comprise the model drivers and are used to compute the transition potential (TP) map—which is central to the calibration of the whole CA-based LUC model (Kolb et al., 2013)— the neighbourhood rules are certainly the most important parameters in these models, as can be seen from the foundational literature (Couclelis, 1985; Gardner, 1970; Tobler, 1979; Ulam, 1952; White and Engelen, 1993). Getting the land-use neighbourhood dynamics (N) right is therefore the most important part of the calibration process. However, it can be extremely challenging to do correctly. In this paper, we present a new approach aimed at facilitating the calibration of neighbourhood influence using an automated rule detection (ARD) procedure.

Specifically, our research aims can be defined as (1) To develop a useful, practical procedure to automate the complex and time-consuming neighbourhood calibration process in CA land-use models; (2) To demonstrate this approach with a simple example, and show how the optimum neighbourhood calibration can be found through a ranking procedure using a wide range of different metrics; (3) To contribute to the theoretical literature on the effect of cell neighbourhood in urban land models (e.g. Díaz-Pacheco et al., 2018; Kocabas and Dragicevic, 2006; Liao et al., 2014; Ménard and Marceau, 2005; Verburg et al., 2004); (4) To share our work in the free and open source "R" software (R Core Team, 2017) so that others can use it in their own research or practice.

For the sake of clarity, we present our results as applied to a very simple urban land-use model, with only urban and non-urban land. But, since we are aware that many model users may be interested in modelling the interaction of multiple land-uses together, we emphasise that the approach we present in this paper is equally applicable to models with multiple-land-uses, e.g. APoLUS (Hewitt et al., 2015). Detailed background to our research, including discussion of the rationale behind the importance of the neighbourhood influence in these types of models is given in the following section.

7.2. Background

LUC driving factors of accessibility and suitability can be easily calibrated using historical LUC data and commonly available geographical information. For instance, accessibility can be calibrated by analysing LUC in various distance classes from existing roads, and suitability can be calibrated by analysing LUC in various categories of terrain slope. However, calibration of the neighbourhood is usually a more complicated procedure involving two key steps: 1) deciding the size of the cell neighbourhood window ($n$size). This is the size of the zone around each cell which is expected to directly influence the state of the land-use in that cell; and 2) deciding the strength of neighbourhood effect itself ($n$rules). Here, $n$rules are user-defined forces of attraction and/or repulsion that decay over the distance covered by the cell neighbourhood window defined in Step 1. The configuration of the neighbourhood setting in CA models of land-use change has been explored previously (Liao et al., 2014; Verburg et al., 2004). However, these studies did not consider both aspects of the neighbourhood (i.e. $n$size and $n$rules) and did not provide a systematic scheme to identify neighbourhood settings which optimise the performance of land-use modelling against multiple spatial landscape metrics (Herold et al., 2005).

For the first case (*n*size), we can see that a large *n*size implies that land-use in one area is affected by land-use both nearby and some distance away, while a small *n*size implies that urban development in one location is influenced only by land-use in adjacent plots. The *n*size is determined by the number of neighbouring cells extending in any single direction, out from the centre, known as the neighbourhood radius (*r*). For example, for *n*size=9 cells (a Moore neighbourhood of 8 cells surrounding a central cell), *r*=1 (see Fig. 7.1). There is no universal agreement on optimal *n*size. It is generally agreed that urban growth in one area may be influenced by the spatial pattern of nearby land-use (Hagoort et al., 2008; Liao et al., 2016; White and Engelen, 2003), implying that a larger *n*size may be needed for accurate simulation. However, other studies have found that increasing *n*size did not significantly affect model results, suggesting that small *n*size may sometimes be adequate for simulating the most important land change tendencies (Díaz-Pacheco et al., 2018). These contrasting findings suggest that each outcome depends on the study area, the model structure, and the input parameters. Therefore, even where researchers have good *a priori* knowledge of the urban area studied, an experimental approach is still likely to be necessary to discover the *n*size that offers the best performance.

For the second case (*n*rules), the shape of the decay curve governs the *strength* of the relationship between near land-uses and those further away. A steep distance decay curve implies that the effect of land-use type of adjacent or nearby grid cells on the grid cell under consideration is much stronger than that for more distant grid cells. A distance decay curve of gentler gradient emphasizes the importance of more distant land. Both *n*size and *n*rules work closely together to determine the spatial pattern of the resulting simulation. A small window with a steep distance decay effect excludes the influence of distant developments entirely. On the other hand, a large window with gentle distance decay effect allows for a greater influence of more distant developments (Fig. 7.1).

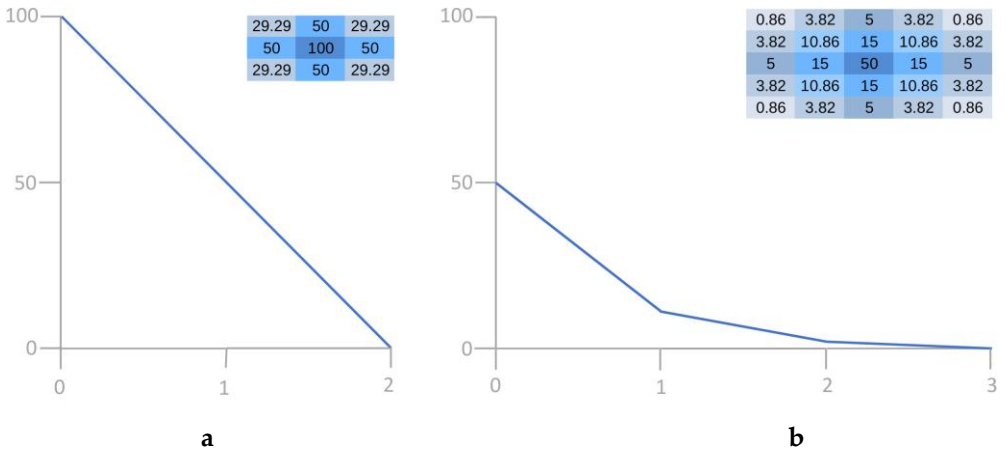


Fig. 7.1. Two examples of neighbourhood dynamics (N). On the left (a: *n*size 9, *r*=1, *n*rules=100,50,0), new urban land is strongly attracted to urban land in locations that are already urban (distance 0), less strongly attracted to neighbouring locations (1 cell distance), with the attraction effect declining to 0 further away (2 cells distance). On the right (*b*: *n*size=25, *r*=2, *n*rules=50,15,5,0), the pattern is the same, but the attraction effect at distance 0 is less strong, and decreases less sharply with distance, reflecting the attraction of urban land to non-urbanised land at greater distances than in *a*.

While the proportional scaling of *n*rules, which determines the shape of the graph, is very important, the numerical values themselves are arbitrary and relate only to the strength of other interactions in the model. In a model with just one land-use, (urban land and non-urban land) the value of *n*rules depends on the relative influence that the modeller wishes to give to the other model drivers (accessibility and suitability) (Fig. 7.2).

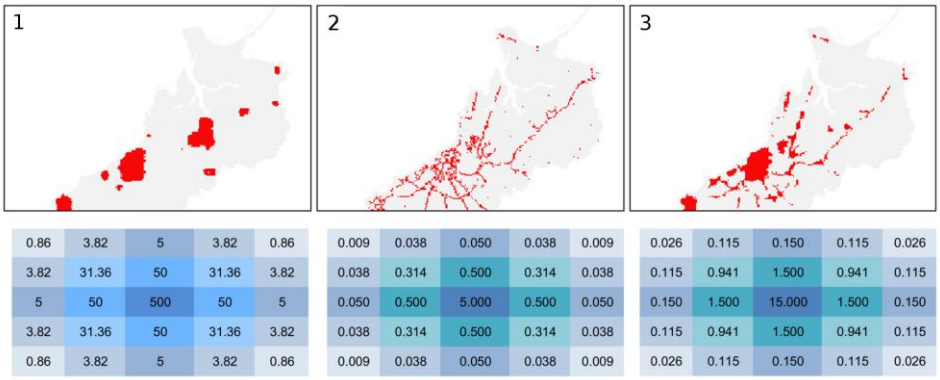


Fig. 7.2. The effect of different *n*rules (bottom) on a simulated urban region. The accessibility, suitability and randomness parameters are held constant in all 3 simulations. Here, *n*size is the same in all simulations. The proportional scaling of *n*rules is the same in all simulations. Simulation 1 (left), the *n*rules values are too strong and overwhelm the effect of accessibility. In Simulation 2, *n*rules are three times as strong as in 3, and new urban land is attracted both to existing urban land and to the road network, producing a more realistic simulation. In Simulation 3, *n*rules values are reduced by a factor of 100; the opposite happens, and accessibility overpowers the *n*rules, so that all new urban land sticks to the road network.

Successful calibration of the neighbourhood setting is thus a time-consuming process characterised by a high degree of uncertainty, which is dependent on multiple factors like the individual characteristics of the case study area as well as the values of other parameters like accessibility and suitability. At the same time, it is also highly dependent on the scale and cell resolution of the model (Díaz-Pacheco et al., 2018). In practice, this uncertainty can only reasonably be reduced by trial and error involving repeated, iterative adjustment and testing of many parameter settings. To facilitate this process, we developed an automatic neighbourhood calibration procedure for the SIMLANDER (Hewitt et al., 2013) land-use modelling environment called *automatic rule detection* (ARD). The proposed ARD tool can be also applied to any similar land-use or urban sprawl models based on cellular automata. The procedure involves repeated testing of different values of *n*size and *n*rules, where the optimum settings are selected by comparing all different outcomes through a ranking approach. We demonstrate our approach by simulating urban growth for the case of Ahvaz city, Iran.

7.3. Methods

7.3.1. Study area

Our study area, Ahvaz city, is the capital and largest city of Khuzestan province located in the southwest of Iran (Fig. 7.3). From 1976 to 2015, the population of Ahvaz increased from 334,399 to 1,338,126 with associated urban growth. The outer parts of the city have great potential for agricultural expansion. The abundance of water and fertile soils has transformed land-use and has generated wealth through agricultural production, which has also increased to supply the expanding local population. Numerous agricultural commodities are produced such as wheat, barley, oilseeds, rice, sugar cane, medicinal herbs, as well as orchard crops such as palm, citrus, and olives (Roodposhti et al., 2018). The abundance of water supplies, rivers, and dams, has also encouraged the fisheries, which are widespread in the area.

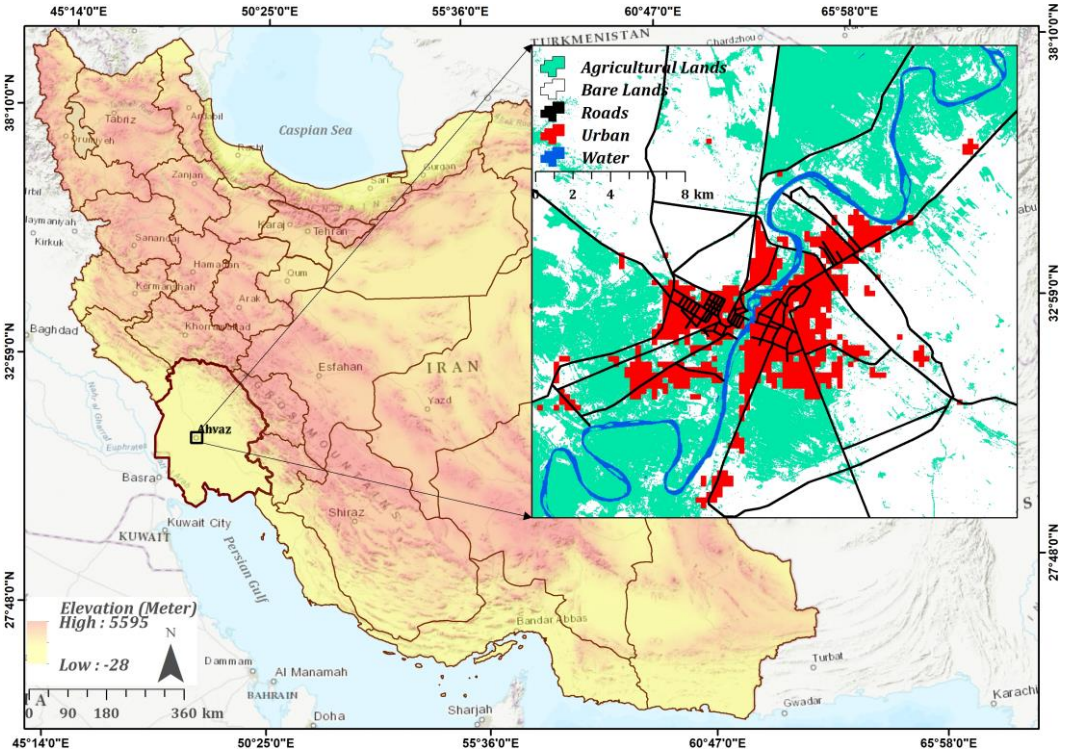


Fig. 7.3. Location and land-use in the study area of Ahvaz, Iran

7.3.2. LUC modelling drivers and data sources

Various groups of independent LUC drivers extracted from three main data sources were selected for LUC simulation. These LUC drivers can be categorised into four groups including neighbourhood dynamics (N), accessibility (A), suitability (S), and randomness (v) (Table 7.1). All LUC drivers were resampled to the 300-metre grid cell resolution of the ENVISAT MERIS sensor land-use products. The study area comprises a 7920-cell raster map with two different land-use classes; urban (value=1) and non-urban (value=0).

Table 7.1. LUC drivers, description, units, and data source applied in this study.

Drivers	Description	Unit	Data source
Land-use 2000	Land-use state (urban/non-urban) of the year 2000	-	ENVISAT
Land-use 2015	Land-use state (urban/non-urban) of the year 2015	-	ENVISAT
Neighbourhood (N)	Number of neighbouring urban cells for each cell	-	Initialised by land-use of 2000
Accessibility (A)	Distance to road networks	metre	Google Earth Engine
Suitability (S)	Slope value	degrees	SRTM DEM
Randomness (v)	Randomness effect by Weibull distribution	-	-

Two land-use maps corresponding to the years 2000 and 2015 were used for the LUC simulation. Following a standard procedure (see, e.g. Hewitt and Díaz-Pacheco 2017), calibration was carried out by applying the method presented in this paper to a land use map of the year 2000 in order to generate a simulation of the year 2015. Fig. 7. shows LUC drivers in Table 7. at the simulation start date of 2000. Note that A and S remain constant throughout the simulation, while neighbourhood and randomness are updated iteratively as the land-use pattern evolves.

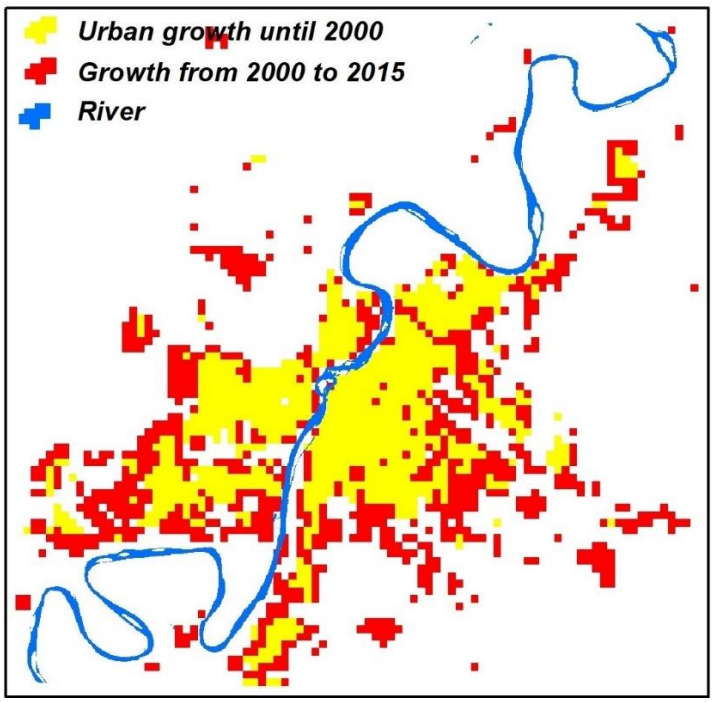


Fig. 7.4. Land use change in the study area, 2000-15.

7.3.3. Overview of the model

Modelling was carried out in the R programming language (R Core Team), and was based on the SIMLANDER R script V.1.0.5 available at https://simlander.wordpress.com/. Apart from the inclusion of the ARD algorithm described in this paper, major modifications applied to SIMLANDER R script V.1.0.5 include, (1) applying a weighting scheme based on Pythagorean theorem for the moving window (Hewitt et al., 2015), and (2) applying fuzzy functions for standardization of selected LUC drivers' components such as distance to roads.

In SIMLANDER, the evolution of the future cell state is determined, as per the conventional CA model, by the following formula, known as the transition potential calculation (Feng et al., 2018b; Hewitt and Díaz-Pacheco, 2017; Wu, 1998):

$$S_{ij}^{t+1} = f(N_{ij}^{t}, A_{ij}^{t}, S_{ij}^{t}, v_{ij}^{t}) \quad (7.1)$$

where, $N_{ij}^{t}, A_{ij}^{t}, S_{ij}^{t}$ and v_{ij}^{t} are a set of the status of neighbouring cells, accessibility, suitability and randomness values at time t, and f is a transition function, often simply the product of all the terms (Hewitt and Díaz-Pacheco, 2017; Hewitt et al., 2014). At each model timestep, all cells in the model are allocated to the locations where the transition potential P_{ij}^{t+1} is highest, on the basis of the cell demand for that timestep, e.g. for a city that grows by 20 cells a year, 120 cells might be allocated at t=1, 140 at t=2, 160 at t=3 and so on. In calibration, the cell demand is determined by dividing the number of new cells of urban land that have emerged between the starting map and the reference map, and the number of years between the two maps. Since the number of cells that will be allocated in the model is controlled by the cell demand, these models are often referred to as "constrained" cellular automata models (White et al., 1997).

Calibration of the model involves replicating observed data as accurately and realistically as possible through the manual adjustment of model parameters (Newland et al., 2018). Typically, N is adjusted first, followed by other parameters, A, S and v, and the model is run to simulate a historic date for which a land-use map is available (in this case the year 2015). In this case, as the goal was testing of

varying settings for N, the *A*, and *S* parameters were calibrated first and held constant throughout the simulation, while neighbourhood and randomness are updated iteratively as the land-use pattern evolves.

To assess the quality of the calibration process, statistical comparison was made between the simulated map and the real date, using a range of standard spatial metrics (see Section 7.3.4.3, Step 5).

7.3.4. Implementation of the approach

Implementation of the ARD tool was characterised by five sequential phases including, (1) data acquisition and preparation, (2) calibration of accessibility, suitability and randomness, (3) automatic calibration of neighbourhood and model goodness-of-fit testing, (4) Prioritisation of optimum neighbourhood settings and (5) evaluation of ARD results obtained from a "*randomised search*" against "*grid search*". Below we explain the implementation approach for each phase.

7.3.4.1. Data acquisition and preparation (ENVI /ArcGIS)

Initially land-use time-series, obtained from the Climate Change Initiative (CCI) land cover archive of the European Space Agency (ESA), which has been claimed to reach 73% accuracy for the 23-class land cover map (Defourny et al., 2009). For A and S parameters, road data and slope data were extracted from Google Earth and the Shuttle Radar Topography Mission (SRTM) digital elevation model, respectively. The road and elevation data were converted to distance to road (i.e. Euclidian) and slope angle using ArcGIS and resampled to the same spatial resolution of ENVISAT land-use maps (i.e. 300m) in ENVI environment. The dataset was then exported to R (R Core Team, 2017) for land-use simulation.

7.3.4.2. Calibration of accessibility, suitability, and randomness

The calibration procedure for accessibility (A), suitability (S), and randomness (v) is explained in detail in the following section. This phase was implemented in R (R Core Team, 2017) using a modified script from the SIMLANDER project (Hewitt et al., 2013). It is emphasized that, while A and S variables are often complex and diverse, including, for example, different types of infrastructure or physical suitability factors such as bedrock geology or soil (see, e.g. Hewitt et al., 2014), we deliberately kept A and S as simple as possible to allow us to focus on the calibration of the neighbourhood parameter N.

Accessibility (A): This is empirically calibrated for each simulated urban growth trajectory through a monotonically decreasing sigmoid function representing distance decay. To this end, the distance-to-road data is applied as an input of the sigmoid function. The sigmoid function is defined as:

$$A_{ij}^{t} = \begin{cases} 1, & if \quad d_{ij} \le a \\ 0.5\left(1 + cos\left(\dfrac{3.14159\left(d_{ij}\text{-}a\right)}{b\text{-}a}\right)\right), & if \quad b > d_{ij} > a \\ 0, & if \quad d_{ij} \ge b \end{cases} \tag{7.2}$$

where d_{ij} is the distance value of the cell a and b are the nearest (300m) and the furthest distance (1500m) to the road network, respectively, and were selected by visual examination of the historical change in urban area from 1995 to 2005.

Suitability (S): the only suitability factor included is the slope angle of the terrain, since Ahvaz is situated predominantly on flat or gently sloping land, and the importance of this variable for urban development is well known from the literature (see, e.g.Clarke et al., 1997; Schneider et al., 2001). To use the slope as a suitability layer, slope values were divided into six classes as shown in Table 7.2.

Table 7.2. Slope angle suitability values applied for LUC model.

Slope angle (degrees)	**Score**
0	1.0
1-5	0.7
5-10	0.5
10-15	0.3
15-20	0.1
20-90	0.0

Randomness (v): the randomness factor (v) applied in this study is provided by a Weibull distribution function that as shown in the following equation (Li and Yeh, 2002; White and Engelen, 1993):

$$v = 1+(-\ln(1-random))\cdot\exp(\alpha) \quad (7.3)$$

where α is a scalable parameter that controls the stochastic effect and *random* refers to a pseudorandom number from the uniform distribution. This can take any desired value, though in practice values between -10 and 1 are usually appropriate, with -10 corresponding to a high stability (almost no randomness) and 1 high volatility (high randomness). A value of 0.5 is usually acceptable for a fairly weak random effect. Values higher than 1 tend to overwhelm the simulations with random transitions. In the other direction, as α tends to -∞, v tends to 1, so values < -10 effectively cancel out the random effect entirely. A completely deterministic model is probably not appropriate for simulating the aggregate effect of human activity in the territory (Hewitt et al., 2014)., so in the usual case values in the range α> -10 <1 are appropriate. For the simulations shown here, α=0.5. In situations where there are many unplanned or disordered land-use transitions, higher values of α may be useful (Barredo et al., 2004).

7.3.4.3. Automatic calibration of N

As described above, the value of neighbourhood (N) for each grid cell at each time-step is a function of the pattern of land-use in the previous time-step under the conditions defined by *n*size and *n*rules. To automate the tedious and time-consuming process of discovering the best-fitting *n*size and *n*rules for simulation of LUC in a given study area, we introduce an automatic rule discovery (ARD) algorithm implemented in R (R Core Team, 2017), which allows neighbourhood to be calculated for each cell according to a five-step procedure described as follows:

***Step 1. Determination of the quantity and maximum size of moving windows*:** the initial requirement for automatic calibration of neighbourhood (N) is to define the "number of moving windows" (i.e. number of n in $\mathbb{N}=\{n_1, n_2, \dots n_n\}$) and their "maximum dimensions" that are required for the automatic calibration process. The former defined how many moving windows will be implemented and tested using ARD while the latter defined the initial dimensions of each constructed window. These values may vary depending on the users' preference and must be entered into the calibration script (Appendix 1) by hand. Any desired number of moving windows may be constructed and applied for calibration. However, in terms of the maximum dimension for the proposed neighbourhood detection tool, the default value is set to 11×11 (i.e. $(2r+1)^2$ where r=5). While larger neighbourhoods (up to 15 x 15, or r=8) can be used in LUC simulation (Engelen et al., 2007; Wickramasuriya et al., 2009), *n*rules are often set to zero for distances beyond 5 cells (van Vliet et al., 2013). In addition, our experimental results also suggest that moving windows bigger than 7×7 are usually less accurate (in terms of both goodness-of-fit and map pattern metrics) than those smaller ones with the same decay rate.

Step 2. Setting up the distance decay/cost functions for the automatic rule detection (ARD) procedure: A spatial cost function forms a raster with a set of values in each cell is the cost for a particular

activity or object in that cell. Here, the cost function is defined by both a decay rate and metric distance from the central cell. For each desired moving window, one cost function will be calculated where each desired cost function has an initial maximum value [5≥ $x_{0,0}$≥50], where distance is equal to "zero", and a unique decay rate. Considering the decay rate for each cost function the interpolation of weights (i.e. w_{ij}) for various cells of a moving window can be calculated using Equation 7.4 (below).

$$w_{xy} = \beta^{(\frac{d_{ij}}{d})} \times x_{0,0} \tag{7.4}$$

where β is the decay rate which is a positive, pseudorandom number from the uniform distribution in the range 0>β,>1. For a randomised search β may be a pseudorandom number from the uniform distribution or any number from a user-defined set for a grid search application. d_{ij} is the distance value of each considered cell from the central cell $x_{0,0}$, d is a constant representing the width of a cell in metres. In equation 4, d depends on the resolution of the model (300 in the present study), and $x_{0,0}$ is the value of central cell (i.e. x=0 and y=0), which is again a pseudorandom number from the uniform distribution in the range 50≥$x_{0,0}$≥5 or an integer number from a user-defined set to be applied for a random or a grid search approach, repectively. Here to properly address the effect of a neighbourhood setting, as a function of $n$size (τ=[2r+1]²) and $n$rules (β), various values of w and β are tested (Fig. 7.5). The random selection approach is faster than a systematic strategy of testing all available possibilities because there is no single optimum rule-set, rather, there are a range of combinations of $n$size and $n$rules that produce acceptable simulations. Nonetheless, ARD is compatible with both randomised and exhaustive grid search approaches. The goal of ARD is to guide the modeller towards the approximate curve shape necessary for their model study case.

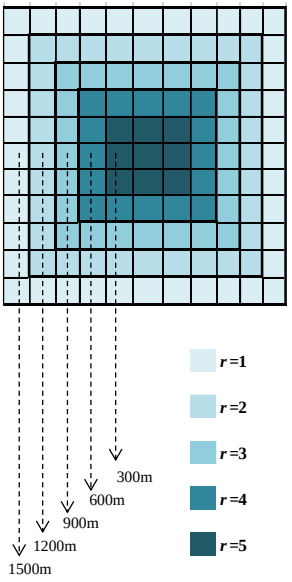


Fig. 7.5. Different sizes of the applied moving windows ($n$size). Here the influence of the cells is a function of the distance from the central cell. The distance of the cells at the corner positions is calculated using the Pythagorean Theorem. Here, cell width is equal to 300m and the default value of maximum r value is set to 6.

***Step 3. Calculation of weighted moving windows*:** A weighting scheme as an inverse function of distance is then applied to every moving window n_1, n_2, … n_n, which is defined by the user. Here, the values of the weighted moving windows for diagonal cells is calculated based on the Pythagorean Theorem. The resultant weighted moving windows will be all the same in terms of dimensions, while various different decay rates are applied.

***Step 4. Applying each constructed moving window for calibration purpose*:** afterwards, each moving window representing a different weighting scheme was applied to the SIMLANDER LUC model in a sequential manner to simulate growth from 2000 to the target year of 2015. The accuracy of the simulation results achieved by each moving window was assessed using a range of statistical

metrics (see Step 5). The process was carried out sequentially, beginning with the largest moving window (11 x 11, *r*=5), reducing the value of *r* each time to obtain successively smaller moving windows until the minimum *r* value of "1" was reached, at which point all values of *n*size had been tested for that rules set (Fig. 7.5). The same procedure was then applied to the next weighted moving window, and so on, until all the different weighting schemes (*n*rules) had been tested for all window sizes (*n*size).

Step 5. Goodness-of-fit testing: In order to determine which of the neighbourhood settings provided a better fit to the reference dataset, simulations produced by each neighbourhood setting were tested using six standard metrics; accuracy, area under the curve (AUC), shape index, fractal dimension, clumpiness and edge density index. These are described as follows:

The *accuracy* (ACC) of the estimation is calculated based on the ratio of hit and correct reject to the total estimations corresponding to the four possible outcomes of the model: *true positive (TP)* (both ground truth and predicted state of a grid cell is "urban"), *false negative (FN)* (only ground-truth state of a grid cell is "urban"), *false positive (FP)* (only predicted state of a grid cell is "urban") and *true negative (TN)* (neither ground truth and predicted state of a grid cell is "urban") are recorded for all simulations (Vaseghi, 2008). Accuracy is calculated as (*TP+TN*/(*TP+FN+FP+TN*)) (Powers, 2011). Though useful as a starting point in goodness-of-fit estimation, accuracy must be used with caution since the value of the score depends on the proportion of positives (*TP+FP*) to negatives (*TN+FN*) in the data. In a *class-imbalanced* dataset, where there are many more positives (negatives) than negatives (positives), results can be very misleading.

To address this problem, the *AUC statistic* (see e.g. Fawcett 2006) uses the relationship between *true positive rates* (*TP*/(*TP+FN*)) and *true negative rates* (*FP*/(*FP+TN*)) to calculate the performance of a prediction model at all decision thresholds in the range 0,1. A model with an AUC of 0 always misses (no cells are simulated correctly and all predictions are false positives), a model with an AUC of 1 always hits (all cells are simulated in the correct location and there are no false positives). An AUC of 0.5 indicates an intermediate state where the model has no discrimination capacity to distinguish between positive class urban and non-urban (Fawcett, 2006b; Feizizadeh et al., 2014c; Shadman Roodposhti et al., 2016). In models where persistence across time steps is built-in, accuracy, AUC and other ratio-based metrics should be adjusted to exclude areas of the model which cannot change (van Vliet et al., 2016). In SIMLANDER, this is not necessary, since all land use is allocated at each time step, so correctly simulating persistence is a key goodness-of-fit indicator.

White (2006) has recommended testing of goodness-of-fit for urban land simulation models on the basis of landscape pattern and structure, rather than pixel-by-pixel metrics like accuracy and AUC. These measures, known collectively as landscape metrics, are familiar to many researchers, especially in the field of landscape ecology, from the FRASTATS package of McGarigal et al. (2002), which is mostly implemented in R in the "*landscapemetrics*" package (Hesselbarth et al., 2019). They are numerical indicators developed for quantifying landscape pattern and work at two fundamental levels, 1) the patch, a cluster of cells separated by other cells or clusters of cells belonging to another land use category, or in this case, non-urban land; 2) the class, which is the land use category that the different patches belong to; in this case, only urban. Since the neighbourhood rules in SIMLANDER control the degree to which cells clump together into patches or disperse along transport corridors (see Fig 7.2), different neighbourhood rules give different values for patch sizes across the class measured. Since there are many metrics, and no consensus on which are the most useful overall (McGarigal, 2014), we selected the metrics to apply from those with demonstrated relevance to goodness-of-fit determination in land use simulations from previous studies. These are shape index, fractal dimension, clumpiness and edge density. Each measure is described briefly described below.

Shape index is a measure of landscape complexity on the basis of comparison against a standard shape (a square) developed by Patton (1975). The shape index of a given class is the ratio between the actual edge length of the class and the hypothetical minimum edge length of the class. The minimum edge length equals the edge length if the class is maximally aggregated (Hesselbarth et al., 2019):

$$SI = \frac{e_i}{\min e_i} \tag{7.5}$$

SI stands for the landscape Shape Index where e_i is the total length of edge (or perimeter) of class *x*, that is the land-use type which is only urban lands in the present case, in terms of the number of cell surfaces and min e_x minimum total length of edge for the same class.

Research on the fractal nature of cities predates CA urban growth models and has been employed as a goodness-of-fit measure in land use simulation modelling by several recent authors (Hewitt and Díaz-Pacheco, 2017; Newland et al., 2015). Mass fractal dimension measures the degree of linearity of the urban class overall, in which plane-filling objects like circles or squares have a value of 2.0 and a line has a value of 1.0.

$$FDI = \frac{2\ln(0.25p_{ij})}{\ln a_{ij}} \tag{7.6}$$

FDI is landscape Fractal Dimension Index where p_{xy} is the perimeter (*m*) of patch *xy* and a_{xy} is the area (m^2) of the same patch. Here, *x* stands for the different patches of each land-use class *y*.

Clumpiness measures the degree of dispersion or aggregation of a class in the range -1 to 1 where -1 is when the patch type is maximally disaggregated to 1 when the patch type is maximally clumped (McGarigal, 2014). Researchers have found this metric to be a useful measure of the degree of fragmentation of a land use category in a land use simulation model (Hewitt et al., 2014; Newland et al., 2015)

$$Given\ G_i = \frac{g_{ij}}{(\sum_{k=1}^{m} g_{ij}) - \min e_i}$$

$$CL = \begin{cases} \dfrac{G_i - P_i}{P_i}\ for\ G_i \prec P_i \,\&\, P_i \prec 0.5\ else \\ \dfrac{G_i - P_i}{1 - P_i} \end{cases} \tag{7.7}$$

Here, g_{xy} is number of like adjacencies (i.e. joins) between pixels of patch type x, that is non-urban lands. CL stands for the clumpiness that equals the proportional deviation of the proportion of like adjacencies involving the corresponding class from that expected under a spatially random distribution. If the proportion of like adjacencies (G_x) is less than the proportion of the landscape comprised of the focal class (P_x) and $P_x < 0.5$, then *CL* equals G_x minus P_x, divided by P_x; else, *CL* equals G_x minus P_x, divided by 1 minus P_x.

Edge density (*ED*) measures the sum of the lengths of all the edges (cells belonging to one class where another class, or, in this case, non-urban land in a given class, is directly adjacent) as a proportion of the total landscape area in the range *ED* >=0, such that *ED*=0 when the landscape is composed of a single patch (Hesselbarth et al 2019). *ED*, like clumpiness, was found to be highly sensitive to variations in CA neighbourhood rules in a recent study (Newland et al., 2015).

$$ED = \frac{edge}{area}(10{,}000) \tag{7.8}$$

ED equals the sum of the lengths (m) of all edge segments in the landscape and *area* is the total landscape area (m^2) multiplied by 10,000 (hectare conversion).

7.3.4.4. Prioritisation of optimum neighbourhood settings

Finally, to allow the modeller to identify the best neighbourhood settings from the range of rule-sets obtained using the ARD tool described above, the different rule-sets were sorted using a ranking procedure. The aim of the ranking procedure was to identify the best and worst neighbourhood settings $\mathbb{N}=\{n_1, n_2, \dots n_n\}$ where each *n* neighbourhood, itself, is composed of various dimensions of $R=\{r_1, r_2, \dots r_m\}$. Here, 50 different settings derived from 10 different 11×11 neighbourhoods are implemented by the ARD tool using randomised search. All these 50 neighbourhoods are then compared using all

measured goodness-of-fit statistics (i.e. AUC, Accuracy, shape index, fractal dimension, clumpiness and edge density index) where the optimum setting is defined by the minimum deviation from the reference values for each goodness-of-fit statistic. To identify the best setting, we first measure absolute deviation from the reference value for every goodness-of-fit statistic (for instance given SI):

$$\Delta_{SI} = \left\{ \left| SI_1 - SI' \right|, \left| SI_2 - SI' \right|, ..., \left| SI_n - SI' \right| \right\} \quad (7.9)$$

where SI' is the reference values of shape index for the ideal simulation, 1 to n, that is measured using the ground truth land-use map (e.g. 2015). We will then achieve Δ_{FDI}, Δ_{CL} and Δ_{ED} in a similar fashion. Afterwards, Δ vectors are ranked (i.e. from 1 to 50) in ascending order using RStudio built-in "*rank*" function (for instance given SI):

$$R_{SI} = rank\left(\Delta_{SI}\right) \quad (7.10)$$

The "*rank*" function returns the sample ranks of the values in a vector *rank*{1,2,3,…*n*}. Ties (i.e., equal values) and missing values can be handled in several ways. For AUC and ACC were ranked using original values while no Δ vectors were required as they are always best at their maximum. Finally, to obtain the ultimate ranking (R_n), the achieved ranking vectors for each goodness-of-fit statistic are passed to a final ranking function:

$$R_n = rank\left(R_{AUC} + R_{ACC} + R_{SI} + R_{FDI} + R_{CL} + R_{ED}\right) \quad (7.11)$$

where R_{AUC}, R_{ACC}, R_{SI}, R_{FDI}, R_{CL} and R_{ED} are ranked vectors for AUC, Accuracy, shape index, fractal dimension, clumpiness and edge density index, respectively. Here, the lower the sum of the *rank* vector for every considered n setting is preferred. Thus, the preferred n setting is the one which has the lowest ranks in R_{AUC}, R_{ACC}, R_{SI}, R_{FDI}, R_{CL} and R_{ED}.

7.3.4.5. Evaluation of ARD using "randomised search" results against "grid search"

In order to evaluate the effectiveness of the applied searching strategy in ARD implementation procedure, we also applied a stepwise "grid search" procedure comprising 250 rule combinations (Table 7.3). That is five times more than the combinations used by the ARD with a randomised search.

Table 7.4. description of three neighbourhood parameters used for grid search implementation.

Parameter	Description	Combinations set	Combinations quantity
$x_{0,0}$	The central cell	X = {5, 10, 15, 20, … ,50}	10
β	Decay rate	B = {x_1/3, x_2/6, x_3/12, x_3/24, x_4/48}	5
r	Neighbourhood radius	R = {*1, 2, 3, 4, 5*}	5

* The total number of combinations applied by grid search is 10×5×5=250.

7.4. Results

In this study, 10 different cost function-based *n*rules were tested by our tool (Fig. 7.6). A steeper slope of the cost-function line equates to a higher decay rate where the cell weight quickly approaches zero as the radius (r) of the neighbourhood increases (i.e. n_{10}); however, a shallower slope line equates to lower decay rates where cell weights are less affected by an increase in distance from the centre. Very low decay rates mean that cell weights are still significant for faraway cells (i.e. n_9).

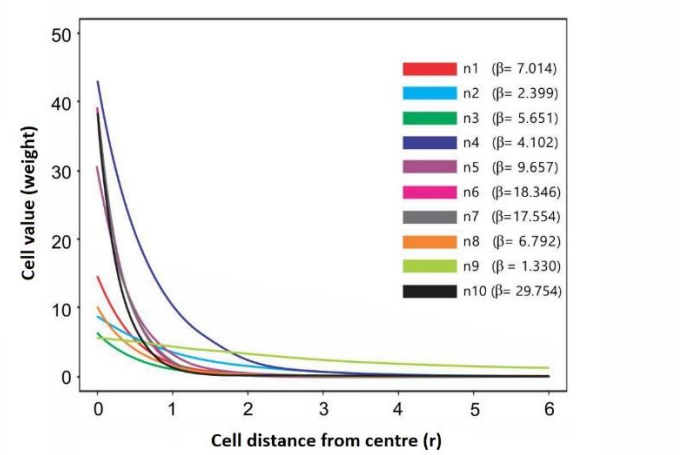


Fig. 7.6. Different cost function applied for automatic implementation of moving windows and their corresponding decay rates (β). These cost functions are results of a randomised grid search that was applied to ARD.

Fig. 7.7 shows a selection of *n*rules (β) corresponding to different decay effects. As the default maximum *n*size(τ) is 11×11 (r=5), in the present application, each neighbourhood window itself contains five combinations of different radii (Fig. 7.).

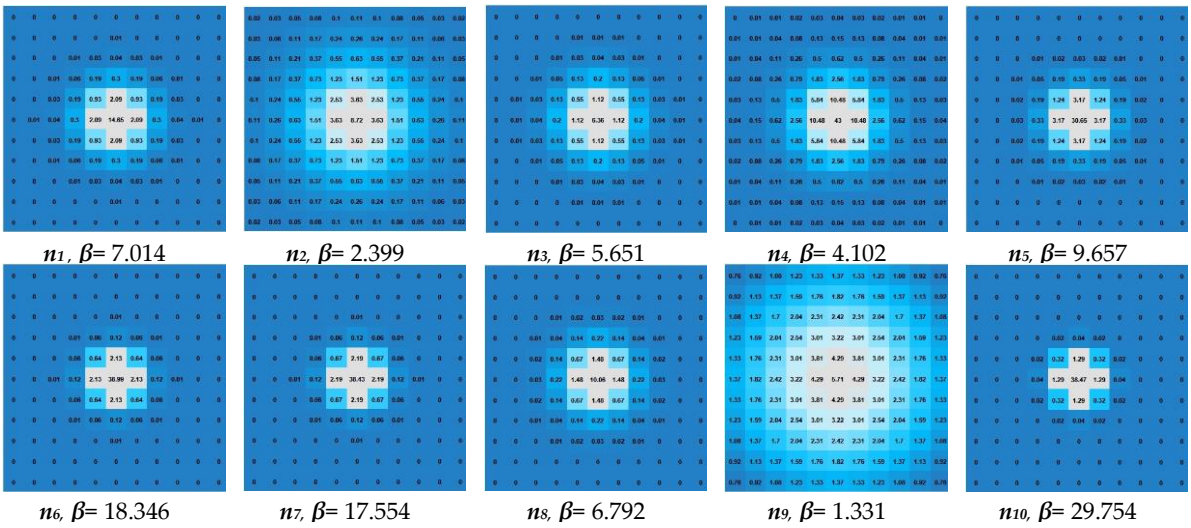


Fig. 7.7. Different settings of *n* characterised by different *n*rules for the default *n*size of 11x11. The numbers refer to the influence value of the cells (*n*), which is a function of the distance from the central cell. The distance of the cells at corner positions is calculated using the Pythagorean Theorem. Each primary moving window is composed of 11×11 cells where four smaller moving windows (i.e. 9×9, 7×7, 5×5 and 3×3) will be derived from the initial moving windows. Each β value refers to decay rate applied for each corresponding *n*rules.

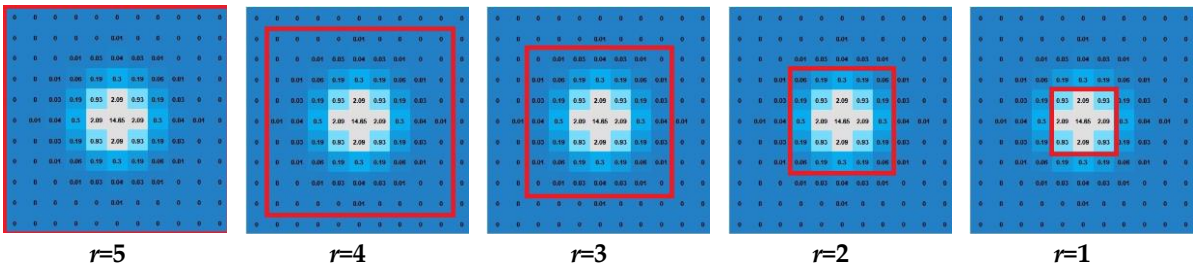


Fig. 7.8. Exemplifying the range of moving window sizes (*n*size) defined by radii *r* tested for the first moving window (n_1). The same approach was followed for all other moving windows.

Results for the neighbourhood settings for 10 best and 10 worst simulations are presented in Table 7.4 and Table 7.5. The Δ values stand for the difference between ground truth and estimated landscape metrics. This is selected from among 50 neighbourhood settings (five simulations for each 11×11 window), which confirms that better simulation results are derived from n_3, n_8 and n_4. It is notable that the 10 best simulations belong to n_3, n_8 and n_4 with moderate values of the decay rate while the majority of the 10 worst simulations belong to n_{10} and n_6 that is implemented using the maximum applied value of the decay rate. The results suggest that neighbourhood size (*n*size) is much less important than neighbourhood rules (*n*rules).

Table 7.4. The achieved simulation metrics of LUC simulation for 10 best applied neighbourhood setting. The AUC and ACC are the best at their maximum values. The ΔSI, ΔFDI, ΔCL, ΔED represent the absolute difference between ideal and estimated values of shape index, fractal dimension index, clumpiness and edge density.

Ideal simulation		*AUC = 1.00*		*ACC = 1.00*		*SI = 8.192*		*FDI = 1.223*		*CL = 0.771*		*ED = 5.853*		-
n-ID	*r-nsize*	*AUC Score*	*AUC rank*	*ACC Score*	*ACC rank*	Δ_{SI} *Score*	*SI rank*	Δ_{FDI} *Score*	*FDI rank*	Δ_{CL} *Score*	*CL rank*	Δ_{ED} *Score*	*ED rank*	**Final rank**
3	4(9×9)	0.821	10	0.879	10	0.783	2	0.010	2	0.024	2	0.542	2	**1st**
8	4(9×9)	0.811	18	0.873	18	0.639	1	0.008	1	0.021	1	0.477	1	**2nd**
3	2(5×5)	0.816	15	0.876	15	0.831	3	0.011	3	0.026	3	0.581	3	**3rd**
3	3(7×7)	0.819	13	0.878	13	0.855	5	0.011	5	0.027	5	0.594	4	**4th**
3	5(11×11)	0.809	21	0.871	21	0.843	4	0.011	4	0.026	4	0.594	4	**5th**
4	1(3×3)	0.811	19	0.872	19	1.072	7	0.015	10	0.034	7	0.754	7	**6th**
8	5(11×11)	0.805	23	0.869	23	0.952	6	0.012	6	0.031	6	0.702	6	**7th**
4	4(9×9)	0.823	8	0.881	8	1.867	14	0.027	15	0.059	14	1.330	14	**8th**
8	2(5×5)	0.807	22	0.870	22	1.096	8	0.014	7	0.036	8	0.810	8	**9th**
4	2(5×5)	0.821	10	0.879	10	2.205	15	0.033	16	0.070	15	1.573	15	**10th**

* The reference values of ideal simulation for the ground truth land-use map of 2015 is shown in brackets.

Table 7.5. The achieved simulation metrics of LUC simulation for 10 worst applied neighbourhood setting. The AUC and ACC are the best at their maximum values. The ΔSI, ΔFDI, ΔCL, ΔED represent the absolute difference between ideal and estimated values of shape index, fractal dimension index, clumpiness and edge density.

Ideal simulation		*AUC = 1.00*		*ACC = 1.00*		*SI = 8.192*		*FDI = 1.223*		*CL = 0.771*		*ED = 5.853*		-
n-ID	*r-nsize*	*AUC Score*	*AUC rank*	*ACC Score*	*ACC rank*	Δ_{SI} *Score*	*SI rank*	Δ_{FDI} *Score*	*FDI rank*	Δ_{CL} *Score*	*CL rank*	Δ_{ED} *Score*	*ED rank*	**Final rank**
10	1(3×3)	0.750	50	0.831	50	9.253	50	0.080	49	0.295	50	6.672	50	**50th**
10	2(5×5)	0.753	48	0.834	48	8.651	49	0.077	48	0.276	49	6.235	49	**49th**
10	4(9×9)	0.753	49	0.833	49	8.253	47	0.074	46	0.263	47	5.949	47	**48th**
10	3(7×7)	0.754	47	0.834	47	8.566	48	0.076	47	0.273	48	6.178	48	**47th**
10	5(11×11)	0.763	41	0.840	41	7.940	46	0.072	44	0.253	46	5.719	46	**46th**
6	5(11×11)	0.755	46	0.835	46	7.494	43	0.069	39	0.239	43	5.403	43	**45th**
6	3(7×7)	0.762	42	0.839	42	7.699	45	0.071	41	0.246	45	5.554	45	**44th**
6	2(5×5)	0.758	45	0.837	45	7.446	42	0.069	38	0.238	42	5.364	42	**43th**
7	1(3×3)	0.764	38	0.841	38	7.639	44	0.070	40	0.244	44	5.511	44	**42th**
6	4(9×9)	0.760	44	0.838	44	7.373	39	0.068	35	0.235	39	5.316	40	**41th**

* The reference values of ideal simulation for the ground truth land-use map of 2015 is shown in brackets.

Using the ARD calibration process described, we found, n_3 with a radius of "4" (i.e. *r*=4) to be the most accurate neighbourhood setting for LUC simulation in the Ahvaz study area. This calibration gave the best overall fit to the reference data set on the basis of cross comparison of multiple metrics using the proposed ranking procedure. On the other hand, n_{10}, with a radius of "1" was the least accurate simulation according to all selected simulation metrics including AUC, ACC, SI, FDI, CL and ED. For the purpose of simplicity and better visualisation, we have presented simulation results only for five most and five least accurate simulation results s (Fig. 7.9 and Fig. 7.10).

Aside from the advantage of simplifying the calibration procedure by removing the need for manual calibration of neighbourhood setting, a key contribution of the proposed tool compared with

the manual calibration process is the fact that the implementation of Pythagorean moving windows using the proposed tool was considerably faster (Table 7.).

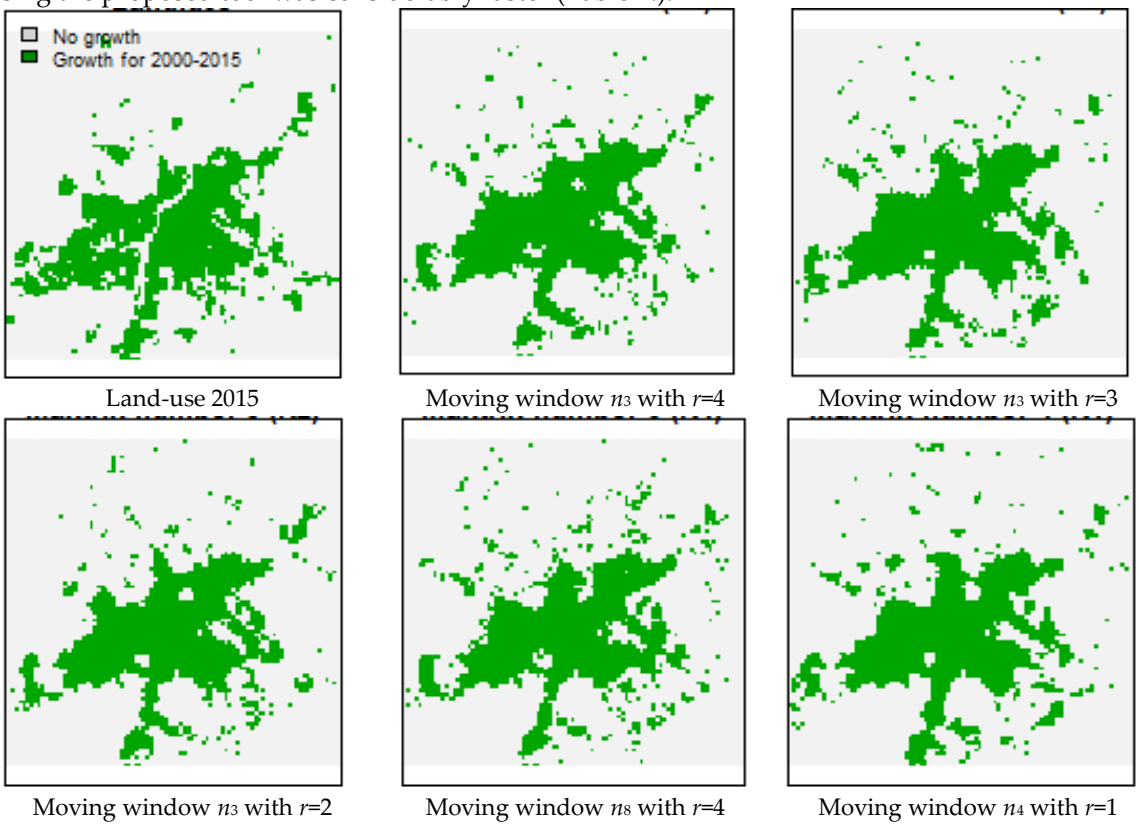


Fig. 7.9. Five best CA-based LUC maps of Ahvaz for the year 2015.

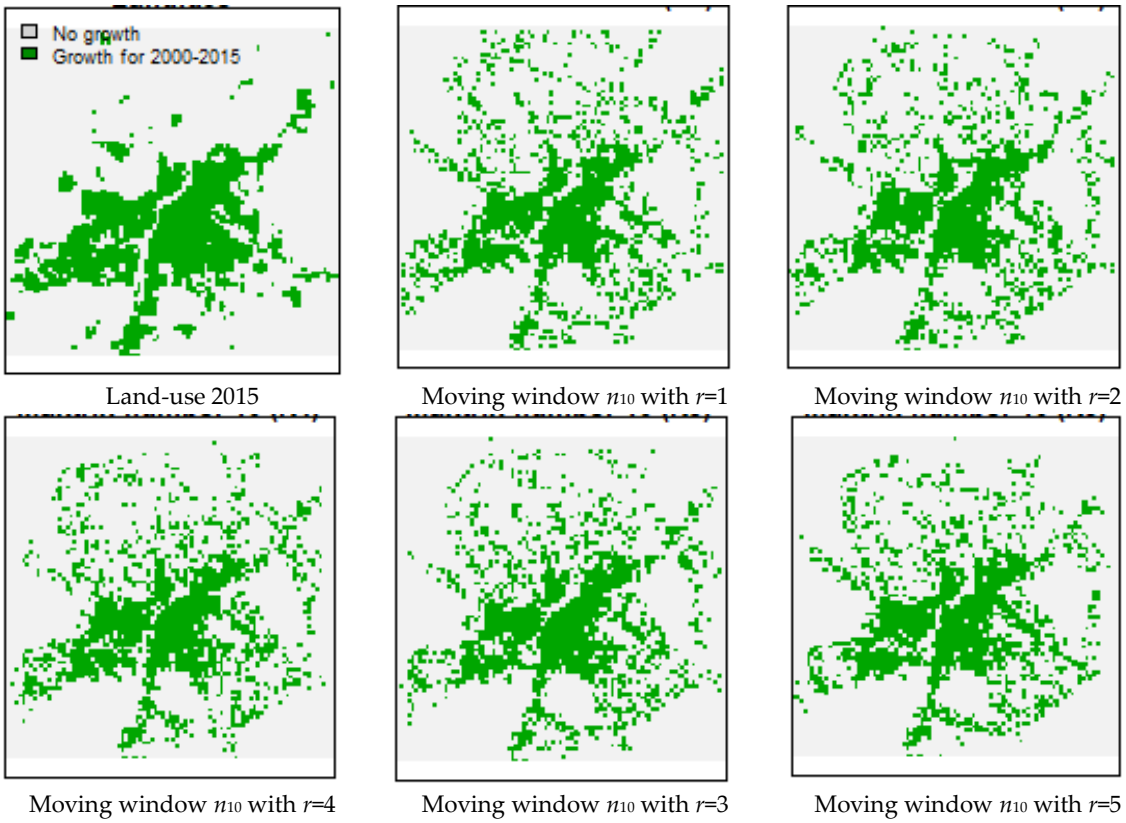


Fig. 7.10. Five worst CA-based LUC maps of Ahvaz for the years 2015.

Table 7.6. Processing time for implementation of neighborhood setting and SIMLANDER simulation and neighborhood setting optimization.

Ideal simulation		*AUC = 1.00*		*ACC = 1.00*		*SI = 8.192*		*FDI = 1.223*		*CL = 0.771*		*ED = 5.853*		-
n-ID	*r-nsize*	*AUC Score*	*AUC rank*	*ACC Score*	*ACC rank*	Δ_{SI} *Score*	*SI rank*	Δ_{FDI} *Score*	*FDI rank*	Δ_{CL} *Score*	*CL rank*	Δ_{ED} *Score*	*ED rank*	**Final rank**
131	5(11×11)	0.814	62	0.875	62	0.0241	1	0.0001	1	0.0002	1	0.0130	1	**1st**

7.5. Discussion

We have presented an automatic neighbourhood detection tool for CA-based LUC models that facilitates and accelerates the calibration process. The procedure described, which we refer to as the Automatic Rule Detection procedure or ARD, allows rapid detection of optimum neighbourhood settings and helps improve understanding of the influence of existing urban land in the growth of new urban areas. In this study, we found that the optimum neighbourhood setting (i.e. N) is more dependent on the distance decay rate (*n*rules) than the neighbourhood dimension (*n*size). Díaz-Pacheco et al. (2018), seeking to explain the apparent lack of variability in calibration results with widely differing rules sets for the case study of Madrid, have suggested that this might have been because more distant land-uses (further than 1 or 2 cells away) had little effect on urban change. This seems logical, in the sense that a building plot might be more likely to be developed when houses are built close by, than if houses are built several streets away.

However, in contrast to these earlier findings, for the city of Ahvaz, we found that the optimum setting of N was highly dependent not only on directly adjacent cells but also on more distant land-use cells, as evidenced by the strong performance of the rules sets with larger moving windows (*r*=4 and 5) (Table 7.4). This suggests that the low variation observed by Díaz Pacheco et al (2018) was due, not because distant land-uses were unimportant, as these authors suggested, but because *n*size was much less important than *n*rules. This is especially likely to be the case there where are only small amounts of urban growth between timesteps, because most new land-use will be allocated to cells immediately adjacent to urban land in the previous time step, and there will be no land to allocate at further distances. In our case, the better-scoring calibrations had moderate distance decay curves, this held true for many different values of *n*size while all 10 worst simulations were derived from those moving windows with highest values of decay rate (β) (Fig. 7.11). In terms of *n*size for different *r* values, no specific pattern was observed (Table 7.4 and Table 7.5).

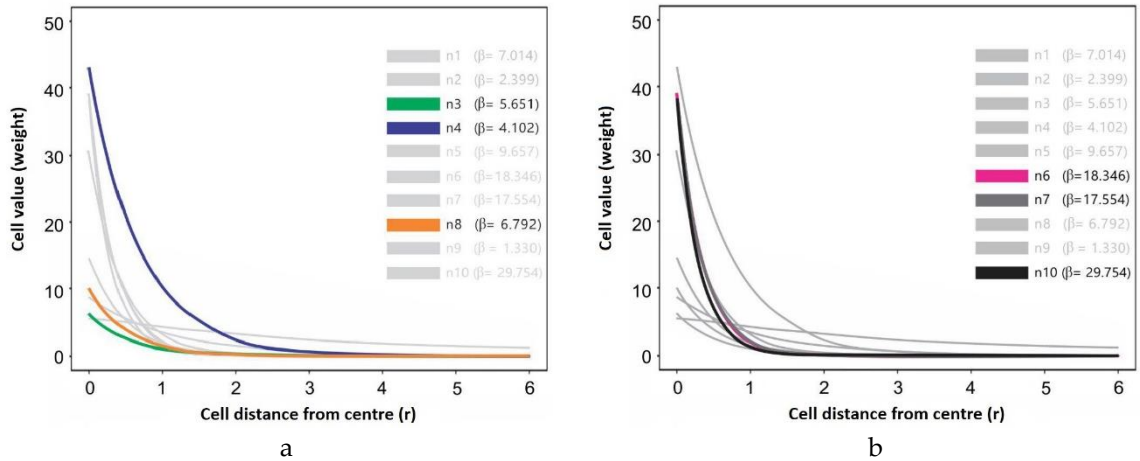


Fig. 7.11. represents (a) the best and (b) worst distance decay curves based on 10 best and 10 worst simulations belonging to n_3, n_8, n_4 and n_{10}, n_6, n_7, respectively. Best curves are highlighted by moderate decay rates while worst curves are characterised by highest values of decay rate (β).

Our results clearly indicate that there is not one, but several, adequate calibration settings for neighbourhood, and that the choice of the "best one" may be subjective. For example, if the researcher

is convinced that distant land-uses are important in their case study area, this *a priori* knowledge should be the most important determinant in selecting the rules set for the calibrated model.

The advantages of our approach over the manual procedure are therefore threefold: 1) the process is much faster, since it does not have to be undertaken by hand; and 2) the process is more thorough, because many more combinations are tested. Thus, even though users may discover an acceptable calibration setting manually by trial-and-error, the ARD procedure allows them to discover if there are better ones, in other words, they are able to optimise the procedure; and 3) the process of goodness-of-fit testing is more rigorous and less arbitrary, since a wide range of spatial metrics are tested rather than just the modeller's favourite ones, which is one key reason for the several different acceptable results. The ARD tool thus provides a way to find rules sets that perform well in terms of statistical goodness-of-fit but that also make intuitive sense to the investigator. These observations highlight the importance of rigorous sensitivity testing of the various combinations of *n*rules and *n*size, and the importance of evaluating simulation accuracy using multiple metrics. In the following sections, we discuss the implications of our work in more detail..

7.5.1. Improving the accuracy of the LUC model through simplification and acceleration of calibration phase

Applying the automatic neighbourhood detection tool for LUC modelling using SIMLANDER is fairly simple in that the only requirement is to select the desired number of moving windows (*n*) and their maximum dimension. This is much faster and more efficient than testing different window sizes and distance decay effects using a manual approach. Moreover, when compared with a systematic grid search approach, the resultant improvement was found to be minor, and came at the expense of a five-fold increase in computation time (Table 7.7).

The implementation of 10 moving windows with maximum *r*-value of "five" (each window is then consist of five windows with different dimensions) for this study was accomplished in less than two seconds (equal to 50 different moving windows), as opposed to manual calibration which can take several days. On this basis, it is clear that automating the discovery of the optimal neighbourhood setting is an important advance in facilitating the calibration of CA models. Following the procedure, we describe is also likely to improve model accuracy in general, by allowing a wider range of neighbourhood settings to be tested more easily. This is especially significant given the importance of neighbourhood (N), compared with other land-use driving factors such as accessibility (A), suitability (S) and randomness (*v*) reported by many studies (Hagoort et al., 2008; Liao et al., 2016; van Vliet et al., 2013).

7.5.2. Achieving a better understanding of optimised neighbourhood setting

Beyond these improvements in model accuracy and speed of calibration, our procedure provides the opportunity to compare the effects of changes in neighbourhood dimensions or the decay rate on the quality of LUC simulation considering all applied metrics such as AUC, ACC, SI, FDI, CL and ED. In this way, model sensitivity analysis is incorporated into the calibration phase, rather than later on, during model testing. In the present study, it was revealed that the impact of changes in decay rate (*n*rules) had a greater impact on simulation results than changes to neighbourhood dimensions (*n*size). In terms of *n*rules, the maximum difference in goodness-of-fit, measured by the AUC value for the least (0.748) and most accurate (0.830) decay rates was more than 0.082. On the other hand, difference of the AUC value for the biggest and smallest moving windows was not more than 0.017. This is also similar for other simulation metrics that are reported in Table 7.4 and Table 7.5, where different neighbourhood dimensions (*n*size) of a decay rate (*n*rules) can remain within ten best or worst simulations.

7.5.3. Towards a better understanding of urban land pattern

It is interesting to note that the five best simulations are all highly aggregated (Fig. 7.9), while the five poorest simulations are very dispersed (Fig. 7.10); Ahvaz, then, is a fairly compact city. The

approach we have described is thus not only appropriate for rapid calibration of urban simulation models, but also a useful classifier of urban form.

In future work, we advocate adopting similar, multiple statistical comparison approaches over the use of just one or two measures for evaluating calibration goodness-of-fit. Though tedious to undertake by hand, the approach we propose here (script provided in Appendix 1), considerably eases this essential task for understanding the pattern and evolution of cities worldwide. The approach we have presented could be further automated in future by testing the neighbourhood effect in combination with a range of accessibility settings.

7.6. Conclusion

CA-based models are effective tools for capturing the complex spatial dynamics of LUC processes. The model calibration is the primary requirement of CA-based LUC models where the neighbourhood effect (N) is always the most crucial parameter in the calibration process. In this study, we have presented and discussed an automatic neighbourhood detection tool for CA-based LUC models, which simplifies and accelerates the discovery of optimum neighbourhood settings, a major requirement of the calibration phase. In addition, the proposed neighbourhood detection tool was also beneficial in achieving a better understanding of urban land-use behaviour as a function of neighbourhood dimension (*n*size) and the decay rate (*n*rules), where the neighbourhood decay rate was found to be more important than neighbourhood size. We also showed how calibration goodness-of-fit can be evaluated by cross comparison of multiple metrics using a ranking procedure. Methodological advances like these can facilitate understanding and implementation of LUC models by reducing the complication and intensive knowledge requirement for model calibration under the popular CA approach.

8. Conclusions

The aim of this research work was to investigate possible strategies to improve quality of spatial models by applying three major types of uncertainty modelling. In this thesis, uncertainty models in their various forms were applied in case study examples such as, landslide modelling, drought sensitivity mapping, land-use change simulation and image processing. To this end, three definitions of uncertainty are reviewed and addressed in this thesis - Type A and B as distinct uncertainties and Type C as fuzzy uncertainty representing an intermediate state of Type A and B. This thesis followed a progressive pathway covering of the main types of uncertainty in spatial sciences that are illustrated by detailed applications. Each chapter addresses one of the thesis objectives (see chapter 1), and has been published or submitted to peer-reviewed journals. Chapter 8, as the concluding chapter, aims to summarise the key contributions of this thesis corresponding the individual research objectives. Firstly, it highlights the usability of developed uncertainty models. Secondly, for each achieved objective, some recommendations and future directions are also provided in this chapter.

8.1. Uncertainty assessment of hyperspectral image classification: deep learning vs random forest

The objective of this chapter, objective 1, was "*to propose and compare uncertainty assessment techniques as a spatial approximator of classification accuracy, which can be used to identify unreliable (with high uncertainty) pixel-level class allocations*". To achieve this objective, two uncertainty assessment techniques that are not dependent on the availability of a test dataset are applied and compared. The proposed uncertainty assessment in this chapter corresponds to Type A and was implemented by calculating the Shannon entropy of class probabilities predicted by emerging deep neural network (DNN) and the popular random forest (RF) algorithm for every pixel. Afterwards, the uncertainty is compared against the classification accuracy of the techniques represented by a modified root mean square error (RMSE). This approach enabled the spatial characterisation of classification accuracy before the validation phase, promoting the assessment of error propagation within the classified imagery products. The achieved results were tested for two different hyperspectral image datasets—Salinas and Indian Pines— and were promising to compensate for the weaknesses of the traditional approaches of map accuracy assessment based on a confusion matrix.

8.2. A robust rule-based ensemble framework using mean-shift segmentation for hyperspectral image classification

The objective of chapter 3, was "t*o develop and implemented a highly accurate and transparent rule-based image classification algorithm based on uncertainty assessment that is derived from probability distribution of possible classification outcomes."*. In order to achieve objective 2, DoTRules—a dictionary of trusted rules—was designed as a supervised rule-based ensemble framework based on mean-shift segmentation and uncertainty assessment, and successfully applied for hyperspectral image classification. The proposed image classification framework was implemented using three world reference hyperspectral image datasets. It was found that the overall accuracy of classification using the proposed ensemble framework was superior to state-of-the-art ensemble algorithms (namely extreme gradient boosting (XGBoost), random forest (RF), rotation forests (RoF), regularised random forest (RRF), as well as two non-ensemble algorithms (i.e. support vector machine (SVM), deep belief network (DBN) and convolutional neural networks (CNN) as the classic deep learning method), at majority of training sample sizes (Please see Table 3.1). DoTRules is both an accurate and transparent approach for image classification. In addition, it is a tool to map rule uncertainty that can be applied as an estimate of classification accuracy prior to image classification.

8.3. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change

The objective of chapter 4, was "*to apply uncertainty assessment -derived from probability distribution of desired outcomes- for implementation of land-use change transition potential maps*". n order to achieve objective 3, we have applied a simplified version of the proposed framework in chapter 4, DoTRules—a dictionary of trusted rules—, as a transparent alternative to calculate transition potential in cellular automata models. Modelling Type A uncertainty was a crucial component of this chapter where Shannon entropy was calculated to assess the uncertainty of each rule, and the most trusted rules were used to project future land-use. In a case study of the Ahvaz region of Iran, the overall accuracy of land-use simulation calibrated using the DoTRules (75.4%) approach was very similar to the random forest (75.8%) simulation results, but DoTRules provides a transparent approach where transition rule information and uncertainty can be readily accessed and interpreted. The results demonstrated the proposed methodology, can provide new insights into Type A uncertainty, and has the potential to improve detection of land-use change processes.

8.4. Drought sensitivity mapping using two one-class support vector machine algorithms

The objective of chapter 5 (i.e. objective 4), was "*to implement a novel drought sensitivity mapping technique through modelling uncertainty levels of decision rules using two one-class support vector machines (OC-SVM)*". Type B uncertainty assessment was addressed in this chapter where a set of fuzzy If-Then rules were implemented to improve the quality of drought sensitivity maps that were developed using two OC-SVM algorithms. The achieved results showed that the proposed methodology based on Type B uncertainty assessment is a promising approach to expose distinct spatio-temporal patterns of drought impacts on vegetation cover, where uncertainty modelling are applied to deal with vagueness of spatial models.

8.5. Fuzzy Shannon entropy: a hybrid GIS-based landslide susceptibility mapping

The objective of chapter 6, was "t*o develop an optimal solution to deal with spatial data uncertainty for both criteria standardisation and prioritisation that is applicable to susceptibility mapping*". This is especially highly beneficial approach when there is insufficient knowledge about a set of spatial criteria belonging to a desired area of interest. For example, in terms of landslide susceptibility, if the mass movement mechanism is not known in the area of interest or if the priority orders of landslide conditioning factors are not clear (Table 6.2), then the proposed methodology is beneficial to compensate for lack of knowledge in landslide susceptibility mapping. To achieve objective 5, I applied a hybrid approach leveraging fuzzy membership functions (FMFs) in combination with Shannon entropy, a well-known information theory-based method. Modelling both Type A and B uncertainty were the crucial components of this chapter for criteria standardisation and prioritisation, respectively. The application domain of this chapter was landslide susceptibility mapping and the role of Type B uncertainty. The performance of the proposed hybrid method with AUC of 0.934 is superior to a multi-criteria evaluation approach using a subjective scheme in this research in comparison with a previous study using the same dataset. through extended fuzzy multi-criteria evaluation with AUC value of 0.894, and was built on the basis of decision makers' evaluation in the same study area. Once again, this highlights the importance of spatial uncertainty for the purpose of modelling and mapping various spatial phenomena such as a landslide.

8.6. Towards automatic calibration of neighbourhood influence in cellular automata land-use models

The objective of chapter 7 (i.e. objective 6), was "*to apply randomness for better calibration of neighbourhood analysis required in automata land-use change models*". Uncertainty assessment Type C was benefitted in this chapter to fulfil the corresponding objective, where a randomized rule detection (RRD) procedure was developed and employed to automate the calibration of neighbourhood setting

as a key requirement of cellular automata (CA) land-use models. Randomisation process here is applied to account for random combination of neighbourhood setting, where β is a random the decay rate which is a positive, pseudorandom number from the uniform distribution in the range $0>\beta,>1$ and directly applies changes in neighbourhood setting. The RRD procedure was useful to facilitate model calibration by allowing rapid identification of the optimum ruleset from a wide range of possible parameter settings, while the goal programming approach facilitates the comparison of simulations using multiple metrics. The RRD approach —centralised around modelling Type C uncertainty — that was presented in chapter 7 demonstrated to be effective in improving simulation accuracy with respect to manual calibration approaches, and increased understanding of neighbourhood dynamics for the urban area studied.

8.7. Recommendations for future studies

This thesis was focused on modelling various types of uncertainty in spatial models. Uncertainty modelling is a fertile ground for future studies aimed to improve the quality of spatial models. For instance, as described in chapter 2, many uncertainty assessment approaches are being developed as a spatial approximator of classification accuracy. This is especially valuable to identify pixel-level class allocations with high uncertainties. Applying different types of uncertainty estimates such as Gini impurity and various forms of entropy may be of researchers' interests in future studies. In addition, referring to chapter 2, the quality of uncertainty assessment that is estimated using various machine-learning algorithms is not the (Figure 2.5 and 2.8). Hence, to provide a superior uncertainty estimate at pixel level different algorithms such as rotation forests (RoF), regularised random forest (RRF), support vector machines (SVM), and other existing state of art algorithms may also be further investigated. The quality of such future studies may be further improved by possible attempts for to improve classification accuracy of each desired algorithm through assessing and minimising uncertainty values such as minimising cross-entropy in deep neural network. Chapters 3 of this thesis is a good example for future studies within the latter category.

Another interesting research topic for future studies may be applying enhanced classification algorithms based on uncertainty assessment in solving spatial problems such as plant population dynamics, forest/bush fire spread, slope failure, debris flow, and urban sprawl etc. Chapter 4 is a practical example for this category in the context of land-use models.

Regardless of possible future studies corresponding to uncertainty Type A as described above, Type B uncertainty, in the form of fuzzy If-Then rules, may be also applied to merge resulting outputs of multiple spatial models and further improve the quality of susceptibility/sensitivity maps. Chapter 5 of this thesis is a good example for this category of possible future studies where the proposed conceptual framework may be generalised to broader susceptibility, sensitivity and risk mapping applications including landslide susceptibility mapping, flood risk mapping etc., where less and more sensitive/susceptible areas can be distinguished using a defined set of spatial criteria. In addition, considering methodological advances in development of machine-learning algorithms one or more state-of-art algorithms (such as XGBoost, DoTRules etc.) can be implemented in the proposed methodology framework, where If-Then fuzzy rules can be applied to integrate the resulting outcomes of each susceptibility class.

Further, such as described in chapter 6, applying various types of uncertainty to improve the quality of spatial models is also another interesting topic for future studies. For instances, Type A and B uncertainty assessment can be applied for criteria standardisation and prioritisation, respectively. This is a good combinational approach to for modelling spatial phenomena where more than one type of uncertainty model is required. Some examples for these kinds of spatial models are modelling land subsidence, earthquake and flood risk mapping but also in multi-hazard risk assessment for a further combination of risk elements. In addition, our approach to identify the priorities of landslide conditioning factors (chapter 6) can be further tested against other approaches such as multi-criteria approaches including analytical hierarchy process (AHP), analytical network process (ANP) etc. Such studies focused on comparing the quality of susceptibility maps where the priority of input predictor

variables are assessed through data-driven (entropy, Gini index etc.) and experts' knowledge (AHP, ANP etc.) will be of considerable contribution to the relevant literature.

Modelling uncertainty Type C for optimisation of neighbourhood setting in LUC models (as explained in chapter 7) may be further pursued by testing our proposed algorithm in different geographic locations. This is a necessary requirement to further improve the quality of RRD for better calibration of neighbourhood analysis applied in automata land-use change models. In addition, applying different optimisation algorithms, rather than simple goal programming is also interesting research topic to improve the quality of neighbourhood optimisation using a random search. In addition, as randomness is also an important component of LUC models (see Eqs. 7.2 and 7.3), further studies on optimising employed random function for LUC models is also an important topic in the context of Type C uncertainty.

Finally, exploring the heterogeneity of the spatial data and their consequential uncertainty may be also helpful in understanding the eventual uncertainties and minimizing them. This is also an important issue to investigate as future topic as it might be more important to look at the reasons underlying the uncertainties and then designing methods to reduce uncertainties.

References

Abdel-Rahman, E.M., Makori, D.M., Landmann, T., Piiroinen, R., Gasim, S., Pellikka, P. and Raina, S.K., 2015. The utility of AISA eagle hyperspectral data and random forest classifier for flower mapping. Remote Sensing, 7(10): 13298-13318.

Acquarelli, J., Marchiori, E., Buydens, L., Tran, T. and van Laarhoven, T., 2017. Convolutional Neural Networks and Data Augmentation for Spectral-Spatial Classification of Hyperspectral Images. arXiv preprint arXiv:1711.05512.

Adep, R.N., Vijayan, A.P., Shetty, A. and Ramesh, H., 2016. Performance evaluation of hyperspectral classification algorithms on AVIRIS mineral data. Perspectives in Science, 8: 722-726.

Adinarayana, J., Gopal Rao, K., Rama Krishna, N., Venkatachalam, P. and Suri, J.K., 1999. A rule-based soil erosion model for a hilly catchment. CATENA, 37(3): 309-318.

Aguilera, F., Valenzuela, L.M. and Botequilha-Leitão, A., 2011. Landscape metrics in the analysis of urban land use patterns: A case study in a Spanish metropolitan area. Landscape and Urban Planning, 99(3-4): 226-238.

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C. and Pradhan, B., 2012a. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Comput. Geosci, 38(1): 23-34.

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C. and Pradhan, B., 2012b. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Computers & Geosciences, 38(1): 23-34.

Akgun, A. and Türk, N., 2010a. Landslide susceptibility mapping for Ayvalik (Western Turkey) and its vicinity by multicriteria decision analysis. Environmental Earth Sciences, 61(3): 595-611.

Akgun, A. and Türk, N., 2010b. Landslide susceptibility mapping for Ayvalik (Western Turkey) and its vicinity by multicriteria decision analysis. Environ. Earth. Sci, 61(3): 595-611.

Al-shalabi, M., Billa, L., Pradhan, B., Mansor, S. and Al-Sharif, A.A.A., 2013. Modelling urban growth evolution and land-use changes using GIS based cellular automata and SLEUTH models: the case of Sana'a metropolitan city, Yemen. Environmental Earth Sciences, 70(1): 425-437.

Al-kheder, S., Wang, J. and Shan, J., 2008. Fuzzy inference guided cellular automata urban-growth modelling using multi-temporal satellite images. International Journal of Geographical Information Science, 22(11-12): 1271-1293.

Allard, D., 2013. J.-P. Chilès, P. Delfiner: Geostatistics: Modeling Spatial Uncertainty. Mathematical Geosciences, 45(3): 377-380.

Amini, M., Abbaspour, K.C. and Johnson, C.A., 2010. A comparison of different rule-based statistical models for modeling geogenic groundwater contamination. Environmental Modelling & Software, 25(12): 1650-1657.

Arsanjani, J.J., Kainz, W. and Mousivand, A.J., 2011. Tracking dynamic land-use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran. International Journal of Image and Data Fusion, 2(4): 329-345.

Atkinson, P.M. and Massari, R., 2011. Autologistic modelling of susceptibility to landsliding in the Central Apennines, Italy. Geomorphology, 130(1): 55-64.

Awad, M., 2014. Sea water chlorophyll-a estimation using hyperspectral images and supervised Artificial Neural Network. Ecological Informatics, 24: 60-68.

Ayalew, L. and Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology, 65(1-2): 15-31.

Ayalew, L., Yamagishi, H. and Ugawa, N., 2004. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. Landslides, 1(1): 73-81.

Aydin, M., 1995. Water key ingredient in Turkish farming, Forum for Applied Research and Public Policy.

Aydin, N.Y., Kentel, E. and Duzgun, H.S., 2013. GIS-based site selection methodology for hybrid renewable energy systems: A case study from western Turkey. Energy Conversion and Management, 70: 90-106.

Aydin, N.Y., Kentel, E. and Duzgun, S., 2010. GIS-based environmental assessment of wind energy systems for spatial planning: A case study from Western Turkey. Renewable and Sustainable Energy Reviews, 14(1): 364-373.

Ayerdi, B. and Graña, M., 2016. Hyperspectral image nonlinear unmixing and reconstruction by ELM regression ensemble. Neurocomputing, 174: 299-309.

Ayerdi, B., Marqués, I. and Graña, M., 2015. Spatially regularized semisupervised Ensembles of Extreme Learning Machines for hyperspectral image segmentation. Neurocomputing, 149: 373-386.

Ballestores Jr, F. and Qiu, Z., 2012. An integrated parcel-based land use change model using cellular automata and decision tree. Proceedings of the International Academy of Ecology and Environmental Sciences, 2(2): 53.

Bannayan, M., Sanjani, S., Alizadeh, A., Lotfabadi, S.S. and Mohamadian, A., 2010. Association between climate indices, aridity index, and rainfed crop yield in northeast of Iran. Field Crops Research, 118(2): 105-114.

Barredo, J., Benavides, A., Hervás, J. and van Westen, C.J., 2000. Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. Int. J. Appl. Earth Obs. Geoinf., 2(1): 9-23.

Barredo, J.I., Demicheli, L., Lavalle, C., Kasanko, M. and McCormick, N., 2004. Modelling Future Urban Scenarios in Developing Countries: An Application Case Study in Lagos, Nigeria. Environment and Planning B: Planning and Design, 31(1): 65-84.

Barredo, J.I., Kasanko, M., McCormick, N. and Lavalle, C., 2003. Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. Landscape and Urban Planning, 64(3): 145-160.

Barreira González, P., Aguilera-Benavente, F. and Gómez-Delgado, M., 2015. Partial validation of cellular automata based model simulations of urban growth: An approach to assessing factor influence using spatial methods. Environmental Modelling & Software, 69: 77-89.

Basse, R.M., Charif, O. and Bódis, K., 2016. Spatial and temporal dimensions of land use change in cross border region of Luxembourg. Development of a hybrid approach integrating GIS, cellular automata and decision learning tree models. Applied Geography, 67: 94-108.

Basse, R.M., Omrani, H., Charif, O., Gerber, P. and Bódis, K., 2014. Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale. Applied Geography, 53: 160-171.

Bathrellos, G.D., Kalivas, D. and Skilodimou, H.D., 2009. GIS-based landslide susceptibility mapping models applied to natural and urban planning in Trikala, Central Greece. Estud. Geol, 65(1): 49-65.

Batty, M. and Xie, Y., 1994. From cells to cities. Environment and planning B: Planning and design, 21(7): S31-S48.

Bauer, T. and Steinnocher, K., 2001. Per-parcel land use classification in urban areas applying a rule-based technique. GeoBIT/GIS, 6: 24-27.

Bazi, Y. and Melgani, F., 2006. Toward an optimal SVM classification system for hyperspectral remote sensing images. IEEE Transactions on geoscience and remote sensing, 44(11): 3374-3385.

Belal, A.-A., El-Ramady, H.R., Mohamed, E.S. and Saleh, A.M., 2014. Drought risk assessment using remote sensing and GIS techniques. Arabian Journal of Geosciences, 7(1): 35-53.

Belgiu, M. and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114: 24-31.

Ben-Hur, A. and Weston, J., 2010. A user's guide to support vector machines. Data mining techniques for the life sciences: 223-239.

Benediktsson, J.A., Garcia, X.C., Waske, B., Chanussot, J., Sveinsson, J.R. and Fauvel, M., 2008. Ensemble methods for classification of hyperspectral data, Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International. IEEE, pp. I-62-I-65.

Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P. and Perrin, C., 2013. Characterising performance of environmental models. Environ. Modell. Softw, 40: 1-20.

Berberoğlu, S., Akın, A. and Clarke, K.C., 2016. Cellular automata modeling approaches to forecast urban growth for adana, Turkey: A comparative approach. Landscape and Urban Planning, 153: 11-27.

Bhuiyan, C., Singh, R. and Kogan, F., 2006. Monitoring drought dynamics in the Aravalli region (India) using different indices based on ground and remote sensing data. International Journal of Applied Earth Observation and Geoinformation, 8(4): 289-302.

Bird, D.K., 2009. The use of questionnaires for acquiring information on public perception of natural hazards and risk mitigation–a review of current knowledge and practice. Natural Hazards and Earth System Sciences, 9(4): 1307-1325.

Boser, B.E., Guyon, I.M. and Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, Proceedings of the fifth annual workshop on Computational learning theory. ACM, pp. 144-152.

Brabb, E.E., **1984**. Innovative approaches to landslide hazard and risk mapping, Proceedings of the fourth international symposium on landslides. Canadian Geotechnical Society, pp. 307–324.

Breiman, L., 1996. Bagging predictors. Machine learning, 24(2): 123-140.

Breiman, L., 2001. Random forests. Machine learning, 45(1): 5-32.

Brereton, R.G. and Lloyd, G.R., 2010. Support vector machines for classification and regression. Analyst, 135(2): 230-267.

Brown, K., Foody, G. and Atkinson, P., 2009. Estimating per-pixel thematic uncertainty in remote sensing classifications. International Journal of Remote Sensing, 30(1): 209-229.

Bruzzone, L. and Prieto, D.F., 2001. Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 39(2): 456-460.

Bryan, B.A., Barry, S. and Marvanek, S., 2009. Agricultural commodity mapping for land use change assessment and environmental management: an application in the Murray–Darling Basin, Australia. Journal of Land Use Science, 4(3): 131-155.

Budimir, M., Atkinson, P. and Lewis, H., 2015. A systematic review of landslide probability mapping using logistic regression. Landslides, 12(3): 419-436.

Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I. and Dick, O.B., 2012. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. Catena, 96: 28-40.

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B. and Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides, 13(2): 361-378.

Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2(2): 121-167.

Burnicki, A.C., 2011. Modeling the probability of misclassification in a map of land cover change. Photogrammetric Engineering & Remote Sensing, 77(1): 39-49.

Burton, I., 1993. The environment as hazard. Guilford Press.

Cancelliere, A., Di Mauro, G., Bonaccorso, B. and Rossi, G., 2007. Drought forecasting using the standardized precipitation index. Water resources management, 21(5): 801-819.

Candel, A., Parmar, V., LeDell, E. and Arora, A., 2015. Deep learning with h2o. H2O.

Candel, A., Parmar, V., LeDell, E. and Arora, A., 2016. Deep learning with H2O. H2O. ai Inc.

Canters, F., 1997. Evaluating the Uncertainty of Area Estimates Derived from Fuuy Land-Cover Classification. Photogrammetric Engineering & Remote Sensing, 63(4): 403-414.

Carrara, A., 1983. Multivariate models for landslide hazard evaluation. J. Int. Ass. Math. Geol, 15(3): 403-426.

Carreira-Perpinán, M.A., 2015. A review of mean-shift algorithms for clustering. arXiv preprint arXiv:1503.00687.

Carter, J.G., 2018. Urban climate change adaptation: Exploring the implications of future land cover scenarios. Cities, 77: 73-80.

Caruana, R., Karampatziakis, N. and Yessenalina, A., 2008. An empirical evaluation of supervised learning in high dimensions, Proceedings of the 25th international conference on Machine learning. ACM, pp. 96-103.

Caruana, R. and Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms, Proceedings of the 23rd international conference on Machine learning. ACM, pp. 161-168.

Castellan, N.J., 2013. Individual and group decision making: current issues. Psychology Press.

Castelvecchi, D., 2016. Can we open the black box of AI? Nature News, 538(7623): 20.

Castilla, A. and Blas, N.G., 2008. Self–organizing map and Cellular automata combined technique for advanced mesh generation in urban and architectural design. International Journal" Information Technologies and knowledge, 2.

Ceamanos, X., Waske, B., Benediktsson, J.A., Chanussot, J., Fauvel, M. and Sveinsson, J.R., 2010. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. International Journal of Image and Data Fusion, 1(4): 293-307.

Chan, J.C.-W. and Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 112(6): 2999-3011.

Chang, N.-B., Parvathinathan, G. and Breeden, J.B., 2008. Combining GIS with fuzzy multicriteria decision-making for landfill siting in a fast-growing urban region. Journal of environmental management, 87(1): 139-153.

Chang, N.-B., Yang, Y.J., Goodrich, J.A. and Daranpob, A., 2010. Development of the Metropolitan Water Availability Index (MWAI) and short-term assessment with multi-scale remote sensing technologies. Journal of environmental management, 91(6): 1397-1413.

Charabi, Y. and Gastli, A., 2011. PV site suitability analysis using GIS-based spatial fuzzy multi-criteria evaluation. Renewable Energy, 36(9): 2554-2561.

Charif, O., Basse, R.-M., Omrani, H. and Trigano, P., 2012. Cellular automata model based on machine learning methods for simulating land use change, Simulation Conference (WSC), Proceedings of the 2012 Winter. IEEE, pp. 1-12.

Chen, F., Wang, K., Van de Voorde, T. and Tang, T.F., 2017. Mapping urban land cover from high spatial resolution hyperspectral data: An approach based on simultaneously unmixing similar pixels with jointly sparse spectral mixture analysis. Remote Sensing of Environment, 196(Supplement C): 324-342.

Chen, T. and Guestrin, C., 2016. Xgboost: A scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, pp. 785-794.

Chen, Y., Lin, Z., Zhao, X., Wang, G. and Gu, Y., 2014. Deep learning-based classification of hyperspectral data. IEEE Journal of Selected topics in applied earth observations and remote sensing, 7(6): 2094-2107.

Chen, Y., Zhao, X. and Jia, X., 2015. Spectral–spatial classification of hyperspectral data based on deep belief network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(6): 2381-2392.

Cheng, J. and Masser, I., 2004. Understanding spatial and temporal processes of urban growth: cellular automata modelling. Environment and Planning B: Planning and Design, 31(2): 167-194.

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. IEEE transactions on pattern analysis and machine intelligence, 17(8): 790-799.

Clarke, K.C., Hoppen, S. and Gaydos, L., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environment and planning B: Planning and design, 24(2): 247-261.

Cohen, J., 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin, 70(4): 213.

Comber, A., Fisher, P., Brunsdon, C. and Khmag, A., 2012. Spatial analysis of remote sensing image classification accuracy. Remote Sensing of Environment, 127: 237-246.

Conforti, M., Pascale, S., Robustelli, G. and Sdao, F., 2014. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena, 113: 236-250.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 37(1): 35-46.

Congalton, R.G. and Green, K., 2008. Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

Conoscenti, C., Di Maggio, C. and Rotigliano, E., 2008. GIS analysis to assess landslide susceptibility in a fluvial basin of NW Sicily (Italy). Geomorphology, 94(3): 325-339.

Constantin, M., Bednarik, M., Jurchescu, M.C. and Vlaicu, M., 2011. Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). Environ. Earth. Sci, 63(2): 397-406.

Couclelis, H., 1985. Cellular worlds: a framework for modeling micro—macro dynamics. Environment and planning A, 17(5): 585-596.

Coulston, J.W., Blinn, C.E., Thomas, V.A. and Wynne, R.H., 2016. Approximating prediction uncertainty for random forest regression models. Photogrammetric Engineering & Remote Sensing, 82(3): 189-197.

Crawford, M.M., Ham, J., Chen, Y. and Ghosh, J., 2003. Random forests of binary hierarchical classifiers for analysis of hyperspectral data, Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on. IEEE, pp. 337-345.

Crosetto, M. and Tarantola, S., 2001. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. International Journal of Geographical Information Science, 15(5): 415-437.

Crosetto, M., Tarantola, S. and Saltelli, A., 2000. Sensitivity and uncertainty analysis in spatial modelling based on GIS. Agriculture, ecosystems & environment, 81(1): 71-79.

Cui, M. and Prasad, S., 2015. Class-dependent sparse representation classifier for robust hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 53(5): 2683-2695.

Cui, Q.-L., Wu, H.-N., Shen, S.-L., Xu, Y.-S. and Ye, G.-L., 2015. Chinese karst geology and measures to prevent geohazards during shield tunnelling in karst region with caves. Natural Hazards, 77(1): 129-152.

D'Ambrosio, D., Gregorio, S.D. and Iovine, G., 2003. Simulating debris flows through a hexagonal cellular automata model: Sciddica s 3–hex. Natural Hazards and Earth System Sciences, 3(6): 545-559.

D'Ambrosio, D., Spataro, W. and Iovine, G., 2006. Parallel genetic algorithms for optimising cellular automata models of natural complex phenomena: an application to debris flows. Computers & Geosciences, 32(7): 861-875.

da Silva, V.d.P.R., Belo Filho, A.F., Almeida, R.S.R., de Holanda, R.M. and da Cunha Campos, J.H.B., 2016. Shannon information entropy for assessing space–time variability of rainfall and streamflow in semiarid region. Science of The Total Environment, 544: 330-338.

Dai, A., 2011. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. Journal of Geophysical Research: Atmospheres, 116(D12).

Dai, A., Trenberth, K.E. and Qian, T., 2004. A global dataset of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming. Journal of Hydrometeorology, 5(6): 1117-1130.

Davis, J. and Blesius, L., 2015. A hybrid physical and maximum-entropy landslide susceptibility model. Entropy, 17(6): 4271-4292.

Davis, T.J. and Keller, C.P., 1997. Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction. International Journal of Geographical Information Science, 11(5): 409-434.

De Luca, A. and Termini, S., 1972. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. Inform.control, 20(4): 301-312.

Defourny, P., Schouten, L., Bartalev, S., Bontemps, S., Caccetta, P., De Wit, A., Di Bella, C.M., Gérard, B., Giri, P. and Gond, V., 2009. Accuracy assessment of a 300 m global land cover map: The GlobCover experience.

Dehghan, H. and Ghassemian, H., 2006. Measurement of uncertainty by the entropy: application to the classification of MSS data. International Journal of Remote Sensing, 27(18): 4005-4014.

Devkota, K.C., Regmi, A.D., Pourghasemi, H.R., Yoshida, K., Pradhan, B., Ryu, I.C., Dhital, M.R. and Althuwaynee, O.F., 2013a. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. Nat. Hazards, 65(1): 135-165.

Devkota, K.C., Regmi, A.D., Pourghasemi, H.R., Yoshida, K., Pradhan, B., Ryu, I.C., Dhital, M.R. and Althuwaynee, O.F., 2013b. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. Natural Hazards, 65(1): 135-165.

Díaz-Pacheco, J., van Delden, H. and Hewitt, R., 2018. The Importance of Scale in Land Use Models: Experiments in Data Conversion, Data Resampling, Resolution and Neighborhood Extent, Geomatic Approaches for Modeling Land Change Scenarios. Springer, pp. 163-186.

Donevska, K.R., Gorsevski, P.V., Jovanovski, M. and Peševski, I., 2012. Regional non-hazardous landfill site selection by integrating fuzzy logic, AHP and geographic information systems. Environmental Earth Sciences, 67(1): 121-131.

Dong, M., Bryan, B.A., Connor, J.D., Nolan, M. and Gao, L., 2015. Land use mapping error introduces strongly-localised, scale-dependent uncertainty into land use and ecosystem services modelling. Ecosystem services, 15: 63-74.

Dou, J., Yamagishi, H., Pourghasemi, H.R., Yunus, A.P., Song, X., Xu, Y. and Zhu, Z., 2015. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Nat. Hazards, 78(3): 1749-1776.

Dutta, R., Das, A. and Aryal, J., 2016. Big data integration shows Australian bush-fire frequency is increasing significantly. Royal Society open science, 3(2): 150241.

Eastman, J.R., 2003. IDRISI Kilimanjaro: guide to GIS and image processing. Clark Labs, Clark University Worcester.

Elshinawy, M.Y., Badawy, A.-H.A., Abdelmageed, W.W. and Chouikha, M., 2010. Comparing one-class and two-class SVM classifiers for normal mammogram detection, Applied Imagery Pattern Recognition Workshop (AIPR), 2010 IEEE 39th. IEEE, pp. 1-7.

Enderton, H. and Enderton, H.B., 2001. A mathematical introduction to logic. Elsevier.

Engelen, G., Lavalle, C., Barredo, J., Van der Meulen, M. and White, R., 2007. The MOLAND modelling framework for urban and regional land-use dynamics, Modelling land-use change. Springer, pp. 297-320.

Ercanoglu, M. and Gokceoglu, C., 2004. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). Eng. Geol, 75(3): 229-250.

Fauvel, M., Benediktsson, J.A., Chanussot, J. and Sveinsson, J.R., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Transactions on Geoscience and Remote Sensing, 46(11): 3804-3814.

Fawcett, T., 2006a. An introduction to ROC analysis. Pattern. Recogn. Lett 27(8): 861-874.

Fawcett, T., 2006b. An introduction to ROC analysis. Pattern recognition letters, 27(8): 861-874.

Feizizadeh, B. and Blaschke, T., 2013. GIS-multicriteria decision analysis for landslide susceptibility mapping: comparing three methods for the Urmia lake basin, Iran. Nat. Hazards, 65(3): 2105-2128.

Feizizadeh, B., Roodposhti, M.S., Jankowski, P. and Blaschke, T., 2014a. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. Comput. Geosci, 73: 208-221.

Feizizadeh, B., Roodposhti, M.S., Jankowski, P. and Blaschke, T., 2014b. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. Computers & geosciences, 73: 208-221.

Feizizadeh, B., Shadman Roodposhti, M., Jankowski, P. and Blaschke, T., 2014c. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. Computers & Geosciences, 73: 208-221.

Felicísimo, Á.M., Cuartero, A., Remondo, J. and Quirós, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides, 10(2): 175-189.

Feng, J., Jiao, L., Liu, F., Sun, T. and Zhang, X., 2016. Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images. Pattern Recognition, 51: 295-309.

Feng, W. and Bao, W., 2017. Weight-Based Rotation Forest for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters, 14(11): 2167-2171.

Feng, Y., Cai, Z., Tong, X., Wang, J., Gao, C., Chen, S. and Lei, Z.J.I.I.J.o.G.-I., 2018a. Urban Growth Modeling and Future Scenario Projection Using Cellular Automata (CA) Models and the R Package Optimx. 7(10): 387.

Feng, Y., Liu, Y. and Tong, X., 2018b. Comparison of metaheuristic cellular automata models: A case study of dynamic land use simulation in the Yangtze River Delta. Computers, Environment and Urban Systems, 70: 138-150.

Feng, Y. and Tong, X., 2019. Incorporation of spatial heterogeneity-weighted neighborhood into cellular automata for dynamic urban growth simulation. GIScience & Remote Sensing: 1-22.

Fisher, P.F., 1999. Models of uncertainty in spatial data. Geographical information systems, 1: 191-205.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote sensing of environment, 80(1): 185-201.

Foody, G.M., Campbell, N., Trodd, N. and Wood, T., 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. Photogrammetric engineering and remote sensing, 58(9): 1335-1341.

Foody, G.M. and Mathur, A., 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. Remote Sensing of Environment, 103(2): 179-189.

Fuglsang, M., Münier, B. and Hansen, H.S., 2013. Modelling land-use effects of future urbanization using cellular automata: An Eastern Danish case. Environmental Modelling & Software, 50: 1-11.

Fukunaga, K. and Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on information theory, 21(1): 32-40.

Galang, J.S., 2004. A comparison of GIS approaches to slope instability zonation in the central Blue Ridge Mountains of Virginia, Virginia Polytechnic Institute and State University.

Gao, L. and Bryan, B.A., 2017. Finding pathways to national-scale land-sector sustainability. Nature, 544(7649): 217-222.

Gao, L., Li, J., Khodadadzadeh, M., Plaza, A., Zhang, B., He, Z. and Yan, H., 2015. Subspace-based support vector machines for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 12(2): 349-353.

Gardner, M., 1970. Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". Scientific American, 223(4): 120-123.

Gbanie, S.P., Tengbe, P.B., Momoh, J.S., Medo, J. and Kabba, V.T.S., 2013. Modelling landfill location using Geographic Information Systems (GIS) and Multi-Criteria Decision Analysis (MCDA): Case study Bo, Southern Sierra Leone. Applied Geography, 36: 3-12.

Gemitzi, A., Petalas, C., Tsihrintzis, V.A. and Pisinaras, V., 2006. Assessment of groundwater vulnerability to pollution: a combination of GIS, fuzzy logic and decision making techniques. Environ. Geol, 49(5): 653-673.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M. and Wolff, E., 2018. Very High Resolution Object-Based Land Use-Land Cover Urban Classification Using Extreme Gradient Boosting. IEEE Geoscience and Remote Sensing Letters.

Giacco, F., Thiel, C., Pugliese, L., Scarpetta, S. and Marinaro, M., 2010. Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs. IEEE Transactions on Geoscience and Remote Sensing, 48(10): 3769-3779.

Gibert, K., Rodríguez-Silva, G. and Rodríguez-Roda, I., 2010. Knowledge discovery with clustering based on rules by states: A water treatment application. Environmental Modelling & Software, 25(6): 712-723.

Gislason, P.O., Benediktsson, J.A. and Sveinsson, J.R., 2004. Random forest classification of multisource remote sensing and geographic data, Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International. IEEE, pp. 1049-1052.

Gislason, P.O., Benediktsson, J.A. and Sveinsson, J.R., 2006. Random forests for land cover classification. Pattern Recognition Letters, 27(4): 294-300.

Goel, P.K., Prasher, S.O., Patel, R.M., Landry, J.A., Bonnell, R.B. and Viau, A.A., 2003. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. Computers and Electronics in Agriculture, 39(2): 67-93.

Golipour, M., Ghassemian, H. and Mirzapour, F., 2016. Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a Bayesian framework. IEEE Transactions on Geoscience and Remote Sensing, 54(2): 805-816.

Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D. and Lee, D.-H., 2015. Challenges in representation learning: A report on three machine learning contests. Neural Networks, 64: 59-63.

Greco, R., Sorriso-Valvo, M. and Catalano, E., 2007. Logistic regression analysis in the evaluation of mass movements susceptibility: the Aspromonte case study, Calabria, Italy. Eng. Geol, 89(1): 47-66.

Guan, D., Li, H., Inohae, T., Su, W., Nagaie, T. and Hokao, K., 2011. Modeling urban land use change by the integration of cellular automaton and Markov model. Ecological Modelling, 222(20–22): 3761-3772.

Gunn, S.R., 1998. Support vector machines for classification and regression. ISIS technical report, 14.

Guo, B., Gunn, S.R., Damper, R.I. and Nelson, J.D., 2006. Band selection for hyperspectral image classification using mutual information. IEEE Geoscience and Remote Sensing Letters, 3(4): 522-526.

Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q., 2017. On calibration of modern neural networks. arXiv preprint arXiv:1706.04599.

Guo, S. and Zhao, H., 2015. Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective. Applied Energy, 158: 390-402.

Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C.P., 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorol. Atmos. Phys, 98(3-4): 239-267.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology, 81(1): 166-184.

Hagoort, M., Geertman, S. and Ottens, H., 2008. Spatial externalities, neighbourhood rules and CA land-use modelling. The Annals of Regional Science, 42(1): 39-56.

Ham, J., Chen, Y., Crawford, M.M. and Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 43(3): 492-501.

Han, H., Yang, C. and Song, J., 2015. Scenario Simulation and the Prediction of Land Use and Land Cover Change in Beijing, China. Sustainability, 7(4): 4260.

Hausser, J., Strimmer, K. and Strimmer, M.K., 2012. Package 'entropy'.

Havenith, H.-B., Strom, A., Caceres, F. and Pirard, E., 2006. Analysis of landslide susceptibility in the Suusamyr region, Tien Shan: statistical and geotechnical approach. Landslides, 3(1): 39-50.

Hayes, M., Svoboda, M., Wall, N. and Widhalm, M., 2011. The Lincoln declaration on drought indices: universal meteorological drought index recommended. Bulletin of the American Meteorological Society, 92(4): 485-488.

He, J. and Kolovos, A., 2018. Bayesian maximum entropy approach and its applications: a review. Stochastic Environmental Research and Risk Assessment, 32(4): 859-877.

Hepinstall, J.A., Marzluff, J.M. and Alberti, M., 2008. Modeling the responses of birds to predicted changes in land cover in an urbanizing region. Models for planning wildlife conservation in large landscapes. Elsevier Science, Burlington, MA, USA: 625-659.

Herold, M., Couclelis, H. and Clarke, K.C., 2005. The role of spatial metrics in the analysis and modeling of urban land use change. Computers, environment and urban systems, 29(4): 369-399.

Hesselbarth, M., Sciaini, M., Nowosad, J. and Hanss, S.J.U.h.r.-s.g.i.l., r package version 0.4, 2019. landscapemetrics: Landscape Metrics for Categorical Map Patterns.

Hewitt, R., de Boer, C., Pacheco, J., Hernández Jiménez, V., Alonso, P., Román, L. and van der Meulen, M., 2015. APoLUS model full system documentation. Project Report for EU FP7 Programme COMPLEX Project (deliverable 3.5) https://www. researchgate. net/publication/292047909_APoLUS_model_full_system_documentation.

Hewitt, R. and Díaz-Pacheco, J., 2017. Stable models for metastable systems? Lessons from sensitivity analysis of a Cellular Automata urban land use model. Computers, Environment and Urban Systems, 62: 113-124.

Hewitt, R., Díaz Pacheco, J. and Moya Gómez, B., 2013. A cellular automata land use model for the R software environment.

Hewitt, R., van Delden, H. and Escobar, F., 2014. Participatory land use modelling, pathways to an integrated approach. Environmental Modelling & Software, 52: 149-165.

Heydari, S.S. and Mountrakis, G., 2018. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. Remote Sensing of Environment, 204: 648-658.

Hijmans, R.J. and van Etten, J., 2014. raster: Geographic data analysis and modeling. R package version, 2(8).

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P. and Sainath, T.N., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6): 82-97.

Hinton, G.E. and Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. science, 313(5786): 504-507.

Hong, H., Pradhan, B., Xu, C. and Bui, D.T., 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena, 133: 266-281.

Hong, Y., Hiura, H., Shino, K., Sassa, K., Suemine, A., Fukuoka, H. and Wang, G., 2005. The influence of intense rainfall on the activity of large-scale crystalline schist landslides in Shikoku Island, Japan. Landslides, 2(2): 97-105.

Houghton, J., Ding, Y., Griggs, D., Noguer, M., Van der Linden, P., Dai, X., Maskell, K. and Johnson, C., 2001. Climate change 2001: the scientific basis–Cambridge University Press. Cambridge.

Hsu, C.-W., Chang, C.-C. and Lin, C.-J., 2003. A practical guide to support vector classification.

Huang, K., Li, S., Kang, X. and Fang, L., 2015. Spectral–Spatial Hyperspectral Image Classification Based on KNN. Sensing and Imaging, 17(1): 1.

Huband, J.M., Bezdek, J.C. and Hathaway, R.J., 2005. bigVAT: Visual assessment of cluster tendency for large data sets. Pattern Recognition, 38(11): 1875-1886.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote sensing of environment, 83(1): 195-213.

Islam, K., Rahman, M.F. and Jashimuddin, M., 2018. Modeling land use change using Cellular Automata and Artificial Neural Network: The case of Chunati Wildlife Sanctuary, Bangladesh. Ecological Indicators, 88: 439-453.

ISO, I. and OIML, B., 1995. Guide to the Expression of Uncertainty in Measurement. Geneva, Switzerland.

Izquierdo-Verdiguier, E., Zurita-Milla, R. and Rolf, A., 2017. On the use of guided regularized random forests to identify crops in smallholder farm fields, Analysis of Multitemporal Remote Sensing Images (MultiTemp), 2017 9th International Workshop on the. IEEE, pp. 1-3.

Jaafari, A., Najafi, A., Pourghasemi, H., Rezaeian, J. and Sattarian, A., 2014. GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. International Journal of Environmental Science and Technology, 11(4): 909-926.

Jain, S.K., Keshri, R., Goswami, A. and Sarkar, A., 2010. Application of meteorological and vegetation indices for evaluation of drought impact: a case study for Rajasthan, India. Natural hazards, 54(3): 643-656.

Jairath, J., 2008. Droughts and integrated water resource management in South Asia: issues, alternatives and futures, 2. SAGE Publications.

Jiang, Z., Huete, A.R., Didan, K. and Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment, 112(10): 3833-3845.

Johnson, C.J. and Gillingham, M.P., 2004. Mapping uncertainty: sensitivity of wildlife habitat ratings to expert opinion. Journal of Applied Ecology, 41(6): 1032-1041.

Jokar Arsanjani, J., 2012. Dynamic Land-Use/Cover Change Simulation: Geosimulation and Multi Agent-Based Modelling. Springer Theses, Springer Verlag.

Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D. and Morisette, J., 2002. An overview of MODIS Land data processing and product status. Remote sensing of Environment, 83(1): 3-15.

Kamusoko, C., Aniya, M., Adi, B. and Manjoro, M., 2009. Rural sustainability under threat in Zimbabwe – Simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. Applied Geography, 29(3): 435-447.

Kamusoko, C. and Gamba, J., 2015. Simulating urban growth using a Random Forest-Cellular Automata (RF-CA) model. ISPRS International Journal of Geo-Information, 4(2): 447-470.

Kanungo, D., Arora, M., Sarkar, S. and Gupta, R., 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Eng. Geol, 85(3): 347-366.

Karavitis, C.A., Alexandris, S., Tsesmelis, D.E. and Athanasopoulos, G., 2011. Application of the standardized precipitation index (SPI) in Greece. Water, 3(3): 787-805.

Kaya, T. and Kahraman, C., 2010. Multicriteria renewable energy planning using an integrated fuzzy VIKOR & AHP methodology: The case of Istanbul. Energy, 35(6): 2517-2527.

Kayabol, K. and Kutluk, S., 2016. Bayesian classification of hyperspectral images using spatially-varying Gaussian mixture model. Digital Signal Processing, 59: 106-114.

Kayastha, P., Dhital, M.R. and De Smedt, F., 2012. Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. Nat. Hazards, 63(2): 479-498.

Keyantash, J. and Dracup, J.A., 2002. The quantification of drought: an evaluation of drought indices. Bulletin of the American Meteorological Society, 83(8): 1167.

Khatami, R., Mountrakis, G. and Stehman, S.V., 2017. Mapping per-pixel predicted accuracy of classified remote sensing images. Remote Sensing of Environment, 191: 156-167.

Khodadadzadeh, M., Li, J., Plaza, A. and Bioucas-Dias, J.M., 2014. A subspace-based multinomial logistic regression for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 11(12): 2105-2109.

Kianisarkaleh, A. and Ghassemian, H., 2016. Nonparametric feature extraction for classification of hyperspectral images with limited training samples. ISPRS Journal of Photogrammetry and Remote Sensing, 119: 64-78.

Kocabas, V. and Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. Computers, Environment and Urban Systems, 30(6): 921-953.

Kolb, M., Mas, J.-F. and Galicia, L., 2013. Evaluating drivers of land-use change and transition potential models in a complex landscape in Southern Mexico. International Journal of Geographical Information Science, 27(9): 1804-1827.

Komac, M., 2006. A landslide susceptibility model using the analytical hierarchy process method and multivariate statistics in perialpine Slovenia. Geomorphology, 74(1): 17-28.

Koomen, E. and Borsboom-van Beurden, J., 2012. Land-use modelling in planning practice. Springer.

Kosko, B., 1986. Fuzzy entropy and conditioning. Inform.Sciences, 40(2): 165-174.

Kritikos, T., Robinson, T.R. and Davies, T.R., 2015. Regional coseismic landslide hazard assessment without historical landslide inventories: A new approach. J. Geophys. Res. Earth Surf., 120(4): 711-729.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, pp. 1097-1105.

Ku, C.-A., 2016. Incorporating spatial regression model into cellular automata for simulating land use change. Applied Geography, 69: 1-9.

Kuhn, M., 2008. Caret package. Journal of statistical software, 28(5): 1-26.

Lagarias, A., 2012. Urban sprawl simulation linking macro-scale processes to micro-dynamics through cellular automata, an application in Thessaloniki, Greece. Applied Geography, 34: 146-160.

Lai, C., Shao, Q., Chen, X., Wang, Z., Zhou, X., Yang, B. and Zhang, L., 2016. Flood risk zoning using a rule mining based on ant colony algorithm. Journal of Hydrology, 542: 268-280.

Lauf, S., Haase, D., Hostert, P., Lakes, T. and Kleinschmit, B., 2012. Uncovering land-use dynamics driven by human decision-making–A combined model approach using cellular automata and system dynamics. Environmental Modelling & Software, 27: 71-82.

Lawrence, R.L., Wood, S.D. and Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). Remote Sensing of Environment, 100(3): 356-362.

Leao, S., Bishop, I. and Evans, D., 2004. Simulating urban growth in a developing nation's region using a cellular automata-based model. Journal of Urban Planning and Development, 130(3): 145-158.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature, 521(7553): 436-444.

Lee, D., Zhou, S., Zhong, X., Niu, Z., Zhou, X. and Zhang, H., 2014. Spatial modeling of the traffic density in cellular networks. IEEE Wireless Communications, 21(1): 80-88.

Lee, S. and Min, K., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. Environ. Geol, 40(9): 1095-1113.

Lee, S. and Pradhan, B., 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides, 4(1): 33-41.

Lee, S., Ryu, J.-H., Won, J.-S. and Park, H.-J., 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Eng. Geol, 71(3): 289-302.

Legleiter, C.J. and Goodchild, M.F., 2005. Alternative representations of in-stream habitat: classification using remote sensing, hydraulic modeling, and fuzzy logic. International Journal of Geographical Information Science, 19(1): 29-50.

Lerner, J.S., Li, Y., Valdesolo, P. and Kassam, K.S., 2015. Emotion and decision making. Psychol, 66.

Li, C., Wang, J., Wang, L., Hu, L. and Gong, P., 2014. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. Remote Sensing, 6(2): 964-983.

Li, J., Chen, F., Cook, E.R., Gou, X. and Zhang, Y., 2007. Drought reconstruction for north central China from tree rings: the value of the Palmer drought severity index. International Journal of Climatology, 27(7): 903-909.

Li, J., Xi, B., Li, Y., Du, Q. and Wang, K., 2018a. Hyperspectral Classification Based on Texture Feature Enhancement and Deep Belief Networks. Remote Sensing, 10(3): 396.

Li, L., Ban, H., Wechsler, S.P. and Xu, B., 2018b. 1.22 - Spatial Data Uncertainty. In: B. Huang (Editor), Comprehensive Geographic Information Systems. Elsevier, Oxford, pp. 313-340.

Li, L., Ge, H. and Gao, J., 2017a. A spectral-spatial kernel-based method for hyperspectral imagery classification. Advances in Space Research, 59(4): 954-967.

Li, L., Zhao, Y., Jiang, D., Zhang, Y., Wang, F., Gonzalez, I., Valentin, E. and Sahli, H., 2013a. Hybrid Deep Neural Network--Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition, Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. IEEE, pp. 312-317.

Li, X., Lao, C., Liu, Y., Liu, X., Chen, Y., Li, S., Ai, B. and He, Z., 2013b. Early warning of illegal development for protected areas by integrating cellular automata with neural networks. Journal of Environmental Management, 130: 106-116.

Li, X., Liu, X. and Gong, P., 2015. Integrating ensemble-urban cellular automata model with an uncertainty map to improve the performance of a single model. International Journal of Geographical Information Science, 29(5): 762-785.

Li, X. and Magill, W., 2001. Modeling fire spread under environmental influence using a cellular automaton approach. Complexity International, 8(1): 1-14.

Li, X. and Yeh, A.G.-O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science, 16(4): 323-343.

Li, Y., Xie, W. and Li, H., 2017b. Hyperspectral image reconstruction by deep convolutional neural network for classification. Pattern Recognition, 63: 371-383.

Liao, J., Tang, L., Shao, G., Qiu, Q., Wang, C., Zheng, S. and Su, X., 2014. A neighbor decay cellular automata approach for simulating urban expansion based on particle swarm intelligence. International Journal of Geographical Information Science, 28(4): 720-738.

Liao, J., Tang, L., Shao, G., Su, X., Chen, D. and Xu, T., 2016. Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations. Environmental Modelling & Software, 75: 163-175.

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. R news, 2(3): 18-22.

Ließ, M., Glaser, B. and Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models. Geoderma, 170: 70-79.

Lioubimtseva, E. and Adams, J., 2004. Possible implications of increased carbon dioxide levels and climate change for desert ecosystems. Environmental Management, 33(1): S388-S404.

Liu, P., Zhang, H. and Eom, K.B., 2017. Active deep learning for classification of hyperspectral images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(2): 712-724.

Liu, X. and Jiang, B., 2011. A novel approach to the identification of urban sprawl patches based on the scaling of geographic space. International Journal of Geomatics and Geosciences, 2(2): 415.

Liu, X., Zhang, J., Ma, D., Bao, Y., Tong, Z. and Liu, X., 2013. Dynamic risk assessment of drought disaster for maize based on integrating multi-sources data in the region of the northwest of Liaoning Province, China. Natural hazards, 65(3): 1393-1409.

Liu, Y., Feng, Y. and Pontius, R.G., 2014. Spatially-explicit simulation of urban growth through self-adaptive genetic algorithm and cellular automata modelling. Land, 3(3): 719-738.

Liucci, L., Melelli, L., Suteanu, C. and Ponziani, F., 2017. The role of topography in the scaling distribution of landslide areas: A cellular automata modeling approach. Geomorphology, 290: 236-249.

Llu, J.G., Mason, P., Hilton, F. and Lee, H., 2004. Detection of rapid erosion in SE Spain: A GIS approach based on ERS SAR coherence imagery: InSAR Application. Photogramm. Eng. Remote Sens, 70(10): 1179-1185.

Loggenberg, K., Strever, A., Greyling, B. and Poona, N., 2018. Modelling water stress in a Shiraz Vineyard using hyperspectral imaging and machine learning. Remote Sensing, 10(2): 202.

Long, Y., Jin, X., Yang, X. and Zhou, Y., 2014. Reconstruction of historical arable land use patterns using constrained cellular automata: A case study of Jiangsu, China. Applied Geography, 52: 67-77.

Loosvelt, L., De Baets, B., Pauwels, V.R. and Verhoest, N.E., 2014. Assessing hydrologic prediction uncertainty resulting from soft land cover classification. Journal of hydrology, 517: 411-424.

Loosvelt, L., Peters, J., Skriver, H., Lievens, H., Van Coillie, F.M., De Baets, B. and Verhoest, N.E., 2012. Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification. International Journal of Applied Earth Observation and Geoinformation, 19: 173-184.

Lopez, S., 2014. Modeling Agricultural Change through Logistic Regression and Cellular Automata: A Case Study on Shifting Cultivation. Journal of Geographic Information System, 06(03): 220-235.

Lotfi, F.H. and Fallahnejad, R., 2010. Imprecise Shannon's entropy and multi attribute decision making. Entropy, 12(1): 53-62.

Löw, F., Michel, U., Dech, S. and Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. ISPRS Journal of Photogrammetry and Remote Sensing, 85: 102-119.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing, 28(5): 823-870.

Lucas, R., Rowlands, A., Brown, A., Keyworth, S. and Bunting, P., 2007. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. ISPRS Journal of photogrammetry and remote sensing, 62(3): 165-185.

Luo, F., Huang, H., Duan, Y., Liu, J. and Liao, Y., 2017a. Local Geometric Structure Feature for Dimensionality Reduction of Hyperspectral Imagery. Remote Sensing, 9(8): 790.

Luo, Y., He, J. and He, Y., 2017b. A rule-based city modeling method for supporting district protective planning. Sustainable Cities and Society, 28: 277-286.

Lv, Q., Niu, X., Dou, Y., Wang, Y., Xu, J. and Zhou, J., 2016. Hyperspectral image classification via kernel extreme learning machine using local receptive fields, Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, pp. 256-260.

Ma, X., Wang, H. and Wang, J., 2016. Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning. ISPRS Journal of Photogrammetry and Remote Sensing, 120: 99-107.

Mahapatra, D., 2014. Analyzing training information from random forests for improved image segmentation. IEEE Transactions on Image Processing, 23(4): 1504-1512.

Majerník, V., 2014. Entropy—A Universal Concept in Sciences. Nat. Sci, 6(7): 552-564.

Makropoulos, C., Butler, D. and Maksimovic, C., 2003. Fuzzy logic spatial decision support system for urban water management. Journal of water resources planning and management, 129(1): 69-77.

Malczewski, J., 2006. Ordered weighted averaging with fuzzy quantifiers: GIS-based multicriteria evaluation for land-use suitability analysis. International journal of applied earth observation and geoinformation, 8(4): 270-277.

Manevitz, L.M. and Yousef, M., 2002. One-class SVMs for document classification. the Journal of machine Learning research, 2: 139-154.

MANR, 2008. Ministry of Agriculture and Natural Resources Tehran, Iran.

Marjanović, M., Kovačević, M., Bajat, B. and Voženílek, V., 2011. Landslide susceptibility assessment using SVM machine learning algorithm. Eng. Geol, 123(3): 225-234.

Martin, A., Laanaya, H. and Arnold-Bos, A., 2006. Evaluation for uncertain image classification and segmentation. Pattern Recognition, 39(11): 1987-1995.

Martinez, G.J., Adamatzky, A. and Alonso-Sanz, R., 2012. Complex dynamics of elementary cellular automata emerging from chaotic rules. International Journal of Bifurcation and Chaos, 22(02): 1250023.

Mas, J.-F., Kolb, M., Paegelow, M., Camacho Olmedo, M.T. and Houet, T., 2014. Inductive pattern-based land use/cover change models: A comparison of four software packages. Environmental Modelling & Software, 51: 94-111.

Mason, P. and Rosenbaum, M., 2002. Geohazard mapping for predicting landslides: an example from the Langhe Hills in Piemonte, NW Italy. Q. J. Eng. Geol. Hydrogeol, 35(4): 317-326.

McGarigal, K., Cushman, S.A., Neel, M.C. and Ene, E., 2002. FRAGSTATS: spatial pattern analysis program for categorical maps.

McGarigal, K.J.D.f.F., 2014. FRAGSTATS help. 4.

McIver, D.K. and Friedl, M.A., 2001. Estimating pixel-scale land cover classification confidence using nonparametric machine learning methods. IEEE Transactions on Geoscience and Remote Sensing, 39(9): 1959-1968.

McKee, T.B., Doesken, N.J. and Kleist, J., 1993. The relationship of drought frequency and duration to time scales, Proceedings of the 8th Conference on Applied Climatology. American Meteorological Society Boston, MA, USA, pp. 179-183.

Ménard, A. and Marceau, D.J., 2005. Exploration of spatial scale sensitivity in geographic cellular automata. Environment and Planning B: Planning and Design, 32(5): 693-714.

MNR, 2010. Landslide event report, Ministry of Natural Resources, Khuzestan Province

Khuzestan, Iran.

Moeinaddini, M., Khorasani, N., Danehkar, A., Darvishsefat, A.A. and zienalyan, M., 2010. Siting MSW landfill using weighted linear combination and analytical hierarchy process (AHP) methodology in GIS environment (case study: Karaj). Waste Management, 30(5): 912-920.

Mon, D.-L., Cheng, C.-H. and Lin, J.-C., 1994. Evaluating weapon system using fuzzy analytic hierarchy process based on entropy weight. Fuzzy. Set. Syst, 62(2): 127-134.

Moore, A.D., Holzworth, D.P., Herrmann, N.I., Brown, H.E., de Voil, P.G., Snow, V.O., Zurcher, E.J. and Huth, N.I., 2014. Modelling the manager: Representing rule-based management in farming systems simulation models. Environmental Modelling & Software, 62: 399-410.

Moreno, N., Wang, F. and Marceau, D.J., 2010. A geographic object-based approach in cellular automata modeling. Photogrammetric Engineering & Remote Sensing, 76(2): 183-191.

Mu, Q., Zhao, M., Kimball, J.S., McDowell, N.G. and Running, S.W., 2013. A remotely sensed global terrestrial drought severity index. Bulletin of the American Meteorological Society, 94(1): 83-98.

Muñoz-Marí, J., Bovolo, F., Gómez-Chova, L., Bruzzone, L. and Camp-Valls, G., 2010. Semisupervised one-class support vector machines for classification of remote sensing data. Geoscience and Remote Sensing, IEEE Transactions on, 48(8): 3188-3197.

Mureriwa, N., Adam, E., Sahu, A. and Tesfamichael, S., 2016. Examining the spectral separability of Prosopis glandulosa from co-existent species using field spectral measurement and guided regularized random forest. Remote Sensing, 8(2): 144.

Mustafa, A., Cools, M., Saadi, I. and Teller, J., 2017. Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM). Land Use Policy, 69: 529-540.

Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M. and Teller, J., 2018a. Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. Computers, Environment and Urban Systems, 67: 147-156.

Mustafa, A., Rienow, A., Saadi, I., Cools, M. and Teller, J., 2018b. Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. European Journal of Remote Sensing, 51(1): 391-401.

Naidoo, L., Cho, M.A., Mathieu, R. and Asner, G., 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. ISPRS Journal of Photogrammetry and Remote Sensing, 69: 167-179.

Nandi, A. and Shakoor, A., 2010. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Eng. Geol, 110(1): 11-20.

Narasimhan, B. and Srinivasan, R., 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. Agricultural and Forest Meteorology, 133(1): 69-88.

Neaupane, K.M. and Piantanakulchai, M., 2006. Analytic network process model for landslide hazard zonation. Eng. Geol, 85(3): 281-294.

Newland, C.P., Maier, H.R. and Newman Van Delden, H., 2015. Relationships between cellular automata based land use models parameters and spatial metrics: enhancing understanding in a calibration context.

Newland, C.P., Maier, H.R., Zecchin, A.C., Newman, J.P. and van Delden, H., 2018. Multi-objective optimisation framework for calibration of Cellular Automata land-use models. Environmental Modelling & Software, 100: 175-200.

Ozdemir, A., 2011. Landslide susceptibility mapping using Bayesian approach in the Sultan Mountains (Akşehir, Turkey). Nat. Hazards, 59(3): 1573-1607.

Pai, D., Sridhar, L., Guhathakurta, P. and Hatwar, H., 2011. District-wide drought climatology of the southwest monsoon season over India based on standardized precipitation index (SPI). Natural hazards, 59(3): 1797-1813.

Pal, M. and Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. IEEE Transactions on Geoscience and Remote Sensing, 48(5): 2297-2307.

Palczewska, A., Palczewski, J., Robinson, R.M. and Neagu, D., 2013. Interpreting random forest models using a feature contribution method, Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on. IEEE, pp. 112-119.

Pan, B., Shi, Z. and Xu, X., 2017. MugNet: Deep learning for hyperspectral image classification using limited samples. ISPRS Journal of Photogrammetry and Remote Sensing.

Pareek, N., Sharma, M.L. and Arora, M.K., 2010. Impact of seismic factors on landslide susceptibility zonation: a case study in part of Indian Himalayas. Landslides, 7(2): 191-201.

Pastor, J., Bonde, J., Johnston, C. and Naiman, R.J., 1991. Markovian analysis of the spatially dependent dynamics of beaver ponds. Lectures on Mathematics in the Life Sciences: 5-28.

Patton, D.R.J.W.S.B., 1975. A diversity index for quantifying habitat" edge". 3(4): 171-173.

Peng, J., Zhang, L. and Li, L., 2016. Regularized set-to-set distance metric learning for hyperspectral image classification. Pattern Recognition Letters, 83: 143-151.

Pontius Jr, R.G. and Millones, M., 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, 32(15): 4407-4429.

Porwollik, V., Müller, C., Elliott, J., Chryssanthacopoulos, J., Iizumi, T., Ray, D.K., Ruane, A.C., Arneth, A., Balkovič, J. and Ciais, P., 2017. Spatial and temporal uncertainty of crop yield aggregations. European Journal of Agronomy, 88: 10-21.

Pourghasemi, H.R., Mohammady, M. and Pradhan, B., 2012a. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena, 97: 71-84.

Pourghasemi, H.R., Pradhan, B. and Gokceoglu, C., 2012b. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Natural Hazards, 63(2): 965-996.

Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

Pozoukidou, G., 2005. Increased usability of urban and land use models. The role of knowledge based systems in facilitating land use forecasting to planning agencies.

Pradhan, B., 2011. Use of GIS-based fuzzy logic relations and its cross application to produce landslide susceptibility maps in three test areas in Malaysia. Environmental Earth Sciences, 63(2): 329-349.

Pradhan, B., 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput. Geosci, 51: 350-365.

Pradhan, B. and Lee, S., 2010. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. Landslides, 7(1): 13-30.

Prasad, M.G. and Arora, M.K., 2014. A simple measure of confidence for fuzzy land-cover classification from remote-sensing data. International journal of remote sensing, 35(24): 8122-8137.

Qiu, F. and Jensen, J., 2004. Opening the black box of neural networks for remote sensing image classification. International Journal of Remote Sensing, 25(9): 1749-1768.

Quartieri, J., Mastorakis, N.E., Iannone, G. and Guarnaccia, C., 2010. A cellular automata model for fire spreading prediction. Latest Trends on Urban Planning and Transportation: 173-178.

Quesney, A., Le Hégarat-Mascle, S., Taconet, O., Vidal-Madjar, D., Wigneron, J., Loumagne, C. and Normand, M., 2000. Estimation of watershed soil moisture index from ERS/SAR data. Remote sensing of environment, 72(3): 290-303.

R Core Team, 2017. R: A Language and Environment for Statistical Computing.

Rangzan, K., Sulaimani, B., Sarsangi, A. and Abshirini, A., 2008. Change detection, mineralogy, desertification mapping in East and Northeast of Ahvaz City, SW Iran using combination of remote sensing methods, GIS and ESAS model. Global Journal of Environmental Research, 2(1): 42-52.

Ray, N. and Burgman, M.A., 2006. Subjective uncertainties in habitat suitability maps. Ecological modelling, 195(3-4): 172-186.

Reshma, R., Sowmya, V. and Soman, K.P., 2016. Dimensionality Reduction Using Band Selection Technique for Kernel Based Hyperspectral Image Classification. Procedia Computer Science, 93: 396-402.

Richards, J. and Jia, X., 1999a. Remote Sensing Digital Image Analysis–Springer. New York.

Richards, J.A. and Jia, X., 1999b. Remote Sensing Digital Image Analysis: An Introduction. Springer-Verlag New York, Inc., 363 pp.

Rienow, A. and Goetzke, R., 2015. Supporting SLEUTH – Enhancing a cellular automaton with support vector machines for urban growth modeling. Computers, Environment and Urban Systems, 49: 66-81.

Rodger, A., Laukamp, C., Haest, M. and Cudahy, T., 2012. A simple quadratic method of absorption feature wavelength estimation in continuum removed spectra. Remote Sensing of Environment, 118(Supplement C): 273-283.

Rojas-Mora, J., Josselin, D., Aryal, J., Mangiavillano, A. and Ellerkamp, P., 2013. The weighted fuzzy barycenter: definition and application to forest fire control in the PACA region. I. J. Agri. Environ. Inform. Sys, 4(4): 48-67.

Rokach, L., 2010. Ensemble-based classifiers. Artificial Intelligence Review, 33(1-2): 1-39.

Roodposhti, M., Rahimi, S. and Beglou, M., 2014a. PROMETHEE II and fuzzy AHP: an enhanced GIS-based landslide susceptibility mapping. Natural Hazards, 73(1): 77-95.

Roodposhti, M.S., Aryal, J. and Bryan, B.A., 2018. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. Environmental Modelling & Software.

Roodposhti, M.S., Aryal, J. and Bryan, B.A., 2019a. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. Environmental Modelling & Software, 112: 70-81.

Roodposhti, M.S., Aryal, J. and Pradhan, B., 2019b. A Novel Rule-Based Approach in Mapping Landslide Susceptibility. Sensors, 19(10): 2274.

Roodposhti, M.S., Rahimi, S. and Beglou, M.J., 2014b. PROMETHEE II and fuzzy AHP: an enhanced GIS-based landslide susceptibility mapping. Natural Hazards, 73(1): 77-95.

Roodposhti, M.S., Rahimi, S. and Beglou, M.J., 2014c. PROMETHEE II and fuzzy AHP: an enhanced GIS-based landslide susceptibility mapping. Nat. Hazards, 73(1): 77-95.

Roodposhti, M.S., Safarrad, T. and Shahabi, H., 2017. Drought sensitivity mapping using two one-class support vector machine algorithms. Atmospheric Research, 193: 73-82.

Royston, S.J., Horsburgh, K.J. and Lawry, J., 2012. Application of rule based methods to predicting storm surge. Continental Shelf Research, 37: 79-91.

Ruff, M. and Czurda, K., 2008. Landslide susceptibility analysis with a heuristic approach in the Eastern Alps (Vorarlberg, Austria). Geomorphology, 94(3-4): 314-324.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1988. Learning representations by back-propagating errors. Cognitive modeling, 5(3): 1.

Russell, S.J., Norvig, P., Canny, J.F., Malik, J.M. and Edwards, D.D., 2003. Artificial intelligence: a modern approach, 2. Prentice hall Upper Saddle River.

Sabokbar, H.F., Roodposhti, M.S. and Tazik, E., 2014. Landslide susceptibility mapping using geographically-weighted principal component analysis. Geomorphology, 226: 15-24.

San Cristóbal, J., 2011. Multi-criteria decision-making in the selection of a renewable energy project in spain: The Vikor method. Renewable energy, 36(2): 498-502.

Santé, I., García, A.M., Miranda, D. and Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. Landscape and Urban Planning, 96(2): 108-122.

Schneider, L.C., Pontius Jr, R.G.J.A., Ecosystems and Environment, 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. 85(1-3): 83-94.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C., 2001. Estimating the support of a high-dimensional distribution. Neural computation, 13(7): 1443-1471.

Seddon, A.W., Macias-Fauria, M., Long, P.R., Benz, D. and Willis, K.J., 2016. Sensitivity of global terrestrial ecosystems to climate variability. Nature, 531(7593): 229-232.

Senf, A., Chen, X.-w. and Zhang, A., 2006. Comparison of one-class SVM and two-class SVM for fold recognition, Neural Information Processing. Springer, pp. 140-149.

Shadman, M., Aryal, J. and Bryan, B., 2017. DoTRules: a novel method for calibrating land-use/cover change models using a Dictionary of Trusted Rules, Proceedings of MODSIM2017, 22nd International Congress on Modelling and Simulation, Hobart, Tasmania, 3-8 December 2017/Syme, G., Hatton MacDonald, D., Fulton, B. and Piantadosi, J.(eds.), pp. 508.

Shadman Roodposhti, M., Aryal, J., Shahabi, H. and Safarrad, T., 2016. Fuzzy shannon entropy: a hybrid GIS-based landslide susceptibility mapping method. Entropy, 18(10): 343.

Shafer, B. and Dezman, L., 1982. Development of a Surface Water Supply Index (SWSI) to assess the severity of drought conditions in snowpack runoff areas, Proceedings of the Western Snow Conference. Colorado State University Fort Collins, CO, pp. 164-175.

Shahabi, H., Ahmad, B. and Khezri, S., 2013. Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: Zab basin). Arab. J. Geosci, 6(10): 3885-3907.

Shahabi, H. and Hashim, M., 2015a. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. Sci. Rep, 5.

Shahabi, H. and Hashim, M., 2015b. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. Scientific reports, 5.

Shahabi, H., Hashim, M. and Ahmad, B.B., 2015. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran. Environ. Earth. Sci, 73(12): 8647-8668.

Shannon, C.E., 2001. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review, 5(1): 3-55.

Shao, Y., Sang, N., Gao, C. and Ma, L., 2017. Probabilistic class structure regularized sparse representation graph for semi-supervised hyperspectral image classification. Pattern Recognition, 63: 102-114.

Silverman, B.W., 2018. Density estimation for statistics and data analysis. Routledge.

Solano, R., Didan, K., Jacobson, A. and Huete, A., 2010. MODIS vegetation index user's guide (MOD13 series). Vegetation index and phenology lab.

Sönmez, F.K., KÖmÜscÜ, A.Ü., Erkan, A. and Turgu, E., 2005. An analysis of spatial and temporal dimension of drought vulnerability in Turkey using the standardized precipitation index. Natural Hazards, 35(2): 243-264.

Srivastava, P., Islam, T., Singh, S., Gupta, M., Petropoulos, G., Gupta, D., Jaafar, W.W. and Prasad, R., 2016. Soil moisture deficit estimation through SMOS soil moisture and MODIS land surface temperature, Satellite Soil Moisture Retrieval. Elsevier, pp. 333-347.

Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment, 62(1): 77-89.

Stein, A., Aryal, J. and Gort, G., 2005a. Use of the Bradley-Terry model to quantify association in remotely sensed images. IEEE transactions on geoscience and remote sensing, 43(4): 852-856.

Stein, A., Aryal, J. and Gort, G., 2005b. Use of the Bradley-Terry model to quantify association in remotely sensed images. IEEE Trans. Geosci. Remote Sens, 43(4): 852-856.

Sujatha, E.R., 2012. Geoinformatics based landslide susceptibility mapping using probabilistic analysis and entropy index of Tevankarai stream sub-watershed, India. Disaster. Adv, 5(3): 26-33.

Tao, Y., Xu, M., Lu, Z. and Zhong, Y., 2018. DenseNet-Based Depth-Width Double Reinforced Deep Learning Neural Network for High-Resolution Remote Sensing Image Per-Pixel Classification. Remote Sensing, 10(5): 779.

Tarabalka, Y., Fauvel, M., Chanussot, J. and Benediktsson, J.A., 2010. SVM-and MRF-based method for accurate classification of hyperspectral images. IEEE Geoscience and Remote Sensing Letters, 7(4): 736-740.

Tax, D.M. and Duin, R.P., 1999. Support vector domain description. Pattern recognition letters, 20(11): 1191-1199.

Taylor, B.N. and Kuyatt, C.E., 1994. Guidelines for evaluating and expressing the uncertainty of NIST measurement results.

Tayyebi, A.H., Tayyebi, A. and Khanna, N., 2014. Assessing uncertainty dimensions in land-use change models: using swap and multiplicative error models for injecting attribute and positional errors in spatial data. International Journal of Remote Sensing, 35(1): 149-170.

Tehrany, M.S., Pradhan, B. and Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of hydrology, 512: 332-343.

Thiery, Y., Malet, J.-P., Sterlacchini, S., Puissant, A. and Maquaire, O., 2007. Landslide susceptibility assessment by bivariate methods at large scales: application to a complex mountainous environment. Geomorphology, 92(1): 38-59.

Tobler, W., 1979. Cellular geography, Philosophy in geography. Springer, pp. 379-386.

Torrens, P.M., 2006. Simulating sprawl. Annals of the Association of American Geographers, 96(2): 248-275.

Toulios, L., Stancalie, G., Struzik, P., Danson, F., Dunkel, Z., Mika, J. and Tsiros, E., 2012. CAPITOLO III POTENTIAL OF REMOTE SENSING TO SUPPORT ThE ASSESSMENT OF CLIMATE ChANGE AND VARIABILITy ON EUROPEAN AGRICULTURE. Climate Change Impacts on Agriculture in Europe: 93.

Tsangaratos, P. and Ilia, I., 2016. Combining fuzzy logic and information theory for producing a landslide susceptibility model, 14th International Congress of the Geological Society of Greece, At Thessaloniki, pp. 34-46.

Tseng, M.-H., Chen, S.-J., Hwang, G.-H. and Shen, M.-Y., 2008. A genetic algorithm rule-based approach for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 63(2): 202-212.

Tsutsumida, N. and Comber, A.J., 2015. Measures of spatio-temporal accuracy for time series land cover data. International Journal of Applied Earth Observation and Geoinformation, 41: 46-55.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote sensing of Environment, 8(2): 127-150.

Ulam, S., 1952. Random processes and transformations, Proceedings of the International Congress on Mathematics. Citeseer, pp. 264-275.

Uslu, F.S., Binol, H., Ilarslan, M. and Bal, A., 2017. Improving SVDD classification performance on hyperspectral images via correlation based ensemble technique. Optics and Lasers in Engineering, 89: 169-177.

Uzuner, Ö., Zhang, X. and Sibanda, T., 2009. Machine learning and rule-based approaches to assertion classification. Journal of the American Medical Informatics Association, 16(1): 109-115.

Valbuena, D., Verburg, P., Bregt, A. and Ligtenberg, A., 2010. An agent-based approach to model land-use change at a regional scale. Landscape Ecology, 25(2): 185-199.

van der Meer, F.D., van der Werff, H.M.A., van Ruitenbeek, F.J.A., Hecker, C.A., Bakker, W.H., Noomen, M.F., van der Meijde, M., Carranza, E.J.M., Smeth, J.B.d. and Woldai, T., 2012. Multi- and hyperspectral geologic remote sensing: A review. International Journal of Applied Earth Observation and Geoinformation, 14(1): 112-128.

van Vliet, J., Bregt, A.K., Brown, D.G., van Delden, H., Heckbert, S. and Verburg, P.H., 2016. A review of current calibration and validation practices in land-change modeling. Environmental Modelling & Software, 82: 174-182.

van Vliet, J., Naus, N., van Lammeren, R.J., Bregt, A.K., Hurkens, J. and van Delden, H., 2013. Measuring the neighbourhood effect to calibrate land use models. Computers, Environment and Urban Systems, 41: 55-64.

Van Vliet, J., White, R. and Dragicevic, S., 2009. Modeling urban growth using a variable grid cellular automaton. Computers, Environment and Urban Systems, 33(1): 35-43.

Van Westen, C. and Getahun, F.L., 2003. Analyzing the evolution of the Tessina landslide using aerial photographs and digital elevation models. Geomorphology, 54(1): 77-89.

Vapnik, V., 2013. The nature of statistical learning theory. Springer science & business media.

Vapnik, V.N. and Vapnik, V., 1998. Statistical learning theory, 1. Wiley New York.

Varnes, D.J., 1978. Slope movement types and processes. Transportation Research Board Special Report(176).

Vaseghi, S.V., 2008. Advanced digital signal processing and noise reduction. John Wiley & Sons.

Vaz, E., Arsanjani, J.J., Phillips, L., Johnson, M., Deener, K., Bonanni, C., Xia, X., Wu, Q., Mou, X. and Lai, Y., 2015. Predicting Urban Growth of the Greater Toronto Area-Coupling a Markov Cellular Automata with Document Meta-Analysis. JOURNAL OF ENVIRONMENTAL INFORMATICS, 25(2): 71-80.

Veldkamp, A. and Lambin, E.F., 2001. Predicting land-use change. Agriculture, ecosystems & environment, 85(1): 1-6.

Verburg, P.H., de Nijs, T.C.M., Ritsema van Eck, J., Visser, H. and de Jong, K., 2004. A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems, 28(6): 667-690.

Vergni, L. and Todisco, F., 2011. Spatio-temporal variability of precipitation, temperature and agricultural drought indices in Central Italy. Agricultural and Forest Meteorology, 151(3): 301-313.

Verstegen, J.A., Karssenberg, D., van der Hilst, F. and Faaij, A.P.C., 2014. Identifying a land use change cellular automaton by Bayesian data assimilation. Environmental Modelling & Software, 53: 121-136.

Vicente-Serrano, S.M., Beguería, S., Lorenzo-Lacruz, J., Camarero, J.J., López-Moreno, J.I., Azorin-Molina, C., Revuelto, J., Morán-Tejeda, E. and Sanchez-Lorenzo, A., 2012. Performance of drought indices for ecological, agricultural, and hydrological applications. Earth Interactions, 16(10): 1-27.

Von Neumann, J. and Burks, A.W., 1966. Theory of self-reproducing automata. IEEE Transactions on Neural Networks, 5(1): 3-14.

Waddell, P., 2002. UrbanSim: Modeling urban development for land use, transportation, and environmental planning. Journal of the American planning association, 68(3): 297-314.

Wakholi, C., Kandpal, L.M., Lee, H., Bae, H., Park, E., S.Kim, M., Mo, C., Lee, W.-H. and Cho, B.-K., 2017. Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics. Sensors and Actuators B: Chemical.

Wang, C. and Shen, H.-W., 2011. Information theory in scientific visualization. Entropy, 13(1): 254-273.

Wang, F., Cao, Y. and Liu, M., 2011. Risk early-warning method for natural disasters based on integration of entropy and DEA model. Appl. Math, 2(01): 23.

Wang, Q. and Shi, W., 2013. Unsupervised classification based on fuzzy c-means with uncertainty analysis. Remote sensing letters, 4(11): 1087-1096.

Wang, W.-D., Guo, J., Fang, L.-G. and Chang, X.-S., 2012. A subjective and objective integrated weighting method for landslides susceptibility mapping based on GIS. Environ.Earth.Sci, 65(6): 1705-1714.

Wei, W., Zhang, J., Zhang, L., Tian, C. and Zhang, Y., 2018. Deep Cube-Pair Network for Hyperspectral Imagery Classification. Remote Sensing, 10(5): 783.

Weng, Q., Quattrochi, D. and Gamba, P.E., 2018. Urban remote sensing. CRC press.

Werick, W., Willeke, G., Guttman, N., Hosking, J. and Wallis, J., 1994. National drought atlas developed. Eos, Transactions American Geophysical Union, 75(8): 89-90.

White, R. and Engelen, G., 1993. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environment and planning A, 25(8): 1175-1199.

White, R. and Engelen, G., 1997. Cellular automata as the basis of integrated dynamic regional modelling. Environment and planning B: Planning and design, 24(2): 235-246.

White, R. and Engelen, G., 2003. A calibration procedure for constrained large neighbourhood cellular automata based land use models, 13th European Colloquium on Theoretical and Quantitative Geography, Lucca, Italy.

White, R., Engelen, G. and Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. Environment and Planning B: Planning and Design, 24(3): 323-343.

White, R.J.J.o.G.S., 2006. Pattern based map comparisons. 8(2): 145-164.

Wickramasuriya, R.C., Bregt, A.K., van Delden, H. and Hagen-Zanker, A., 2009. The dynamics of shifting cultivation captured in an extended Constrained Cellular Automata land use model. Ecological Modelling, 220(18): 2302-2309.

Wilhite, D.A., 2012. Drought Assessment, Management, and Planning: Theory and Case Studies: Theory and Case Studies, 2. Springer Science & Business Media.

Wolfram, S., 2002. A new kind of science, 5. Wolfram media Champaign.

Wu, F., 1998. Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. Journal of Environmental Management, 53(4): 293-308.

Xia, J., Chanussot, J., Du, P. and He, X., 2015. Spectral–spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields. IEEE Transactions on Geoscience and Remote Sensing, 53(5): 2532-2546.

Xia, J., Du, P., He, X. and Chanussot, J., 2014. Hyperspectral remote sensing image classification based on rotation forest. IEEE Geoscience and Remote Sensing Letters, 11(1): 239-243.

Xia, J., Falco, N., Benediktsson, J.A., Chanussot, J. and Du, P., 2016. Class-separation-based rotation forest for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 13(4): 584-588.

Xia, J., Ghamisi, P., Yokoya, N. and Iwasaki, A., 2018. Random Forest Ensembles and Extended Multiextinction Profiles for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing, 56(1): 202-216.

Xian, G. and Crane, M., 2005. Assessments of urban growth in the Tampa Bay watershed using remote sensing data. Remote Sensing of Environment, 97(2): 203-215.

Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y. and Huang, Z., 2006. Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. Landscape and urban planning, 75(1-2): 69-80.

Xie, Y., Sha, Z. and Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. Journal of plant ecology, 1(1): 9-23.

Xu, C., Dai, F., Xu, X. and Lee, Y.H., 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. Geomorphology, 145: 70-80.

Xu, J., Gu, B., Guo, Y., Chang, J., Ge, Y., Min, Y. and Jin, X., 2010. A cellular automata model for population dynamics simulation of two plant species with different life strategies, Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on. IEEE, pp. 517-523.

Yalcin, A., 2008. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. CATENA, 72(1): 1-12.

Yang, C., Tan, Y., Bruzzone, L., Lu, L. and Guan, R., 2017. Discriminative Feature Metric Learning in the Affinity Propagation Model for Band Selection in Hyperspectral Images. Remote Sensing, 9(8): 782.

Yang, J., Su, J., Chen, F., Xie, P. and Ge, Q., 2016. A local land use competition cellular automata model and its application. ISPRS International Journal of Geo-Information, 5(7): 106.

Yang, Q., Li, X. and Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. Computers & geosciences, 34(6): 592-602.

Yang, X., Chen, L., Li, Y., Xi, W. and Chen, L., 2015. Rule-based land use/land cover classification in coastal areas using seasonal remote sensing imagery: a case study from Lianyungang City, China. Environmental Monitoring and Assessment, 187(7): 449.

Yang, X., Zheng, X.-Q. and Chen, R., 2014. A land use change model: Integrating landscape pattern indexes and Markov-CA. Ecological Modelling, 283(Supplement C): 1-7.

Ye, S., Pontius, R.G. and Rakshit, R., 2018. A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. ISPRS Journal of Photogrammetry and Remote Sensing, 141: 137-147.

Yeon, Y.-K., Han, J.-G. and Ryu, K.H., 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. Eng. Geol., 116(3): 274-283.

Youssef, A.M., Maerz, N.H. and Hassan, A.M., 2009. Remote sensing applications to geological problems in Egypt: case study, slope instability investigation, Sharm El-Sheikh/Ras-Nasrani Area, Southern Sinai. Landslides, 6(4): 353-360.

Youssef, A.M., Pourghasemi, H.R., El-Haddad, B.A. and Dhahry, B.K., 2016. Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia. Bull. Eng. Geol. Environ, 75(1): 63-87.

Yu, Q., Gong, P., Tian, Y.Q., Pu, R. and Yang, J., 2008. Factors affecting spatial variation of classification uncertainty in an image object-based vegetation mapping. Photogrammetric Engineering & Remote Sensing, 74(8): 1007-1018.

Yu, S., Jia, S. and Xu, C., 2017. Convolutional neural networks for hyperspectral image classification. Neurocomputing, 219: 88-98.

Yufeng, S. and Fengxiang, J., 2009. Landslide stability analysis based on generalized information entropy, Environmental Science and Information Application Technology, 2009. ESIAT 2009. International Conference on. IEEE, pp. 83-85.

Zadeh, L.A., 1965a. Fuzzy sets. Information and control, 8(3): 338-353.

Zadeh, L.A., 1965b. Information and control. Fuzzy sets., 8(3): 338-353.

Zadeh, L.A., 1976. A fuzzy-algorithmic approach to the definition of complex or imprecise concepts, Systems Theory in the Social Sciences. Springer, pp. 202-282.

Zarco-Tejada, P.J., González-Dugo, M.V. and Fereres, E., 2016. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. Remote Sensing of Environment, 179(Supplement C): 89-103.

Zhang, L., Zhang, L. and Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine, 4(2): 22-40.

Zhang, Q., Xu, C.-Y. and Zhang, Z., 2009. Observed changes of drought/wetness episodes in the Pearl River basin, China, using the standardized precipitation index and aridity index. Theoretical and Applied Climatology, 98(1-2): 89-99.

Zhang, X., Tong, X. and Liu, M., 2008. Genetic algorithm optimized neutral network classification and overlapping pixel change detection based on remote sensing for urban sprawl: a case study in Jiading district of Shanghai, China, Proceedings of the ISPRS XXIst Congress, July, pp. 3-11.

Zheng, Z., Huang, W., Li, S. and Zeng, Y., 2017. Forest fire spread simulating model using cellular automaton with extreme learning machine. Ecological Modelling, 348: 33-43.

Zhong, Y. and Zhang, L., 2012. An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 50(3): 894-909.

Zhu, D., Yao, H., Jiang, B. and Yu, P., 2018. Negative Log Likelihood Ratio Loss for Deep Neural Network Classification. arXiv preprint arXiv:1804.10690.

Zimmermann, H.J., 2010. Fuzzy set theory. Wiley Interdisciplinary Reviews: Computational Statistics, 2(3): 317-332.

Zongji, Y., Jianping, Q. and Xiaogang, Z., 2010. Regional landslide zonation based on entropy method in Three Gorges area, China, Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on. IEEE, pp. 1336-1339.

Zou, Q., Zhou, J., Zhou, C., Song, L. and Guo, J., 2013. Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. Stochastic Environmental Research and Risk Assessment, 27(2): 525-546.